

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 17, 2002, 19:22:50 : Search time 2662 Seconds
(without alignments)
16815.125 Million cell updates/sec

Title: US-10-008-355-1
Perfect score: 2139
Sequence: 1 atgcaaatgaattaaaaag.....aagagctgaagtgtactctaa 2139

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*\n1: gb_ba:*\n2: gb_htg:*\n3: gb_in:*\n4: gb_om:*\n5: gb_ov:*\n6: gb_pat:*\n7: gb_ph:*\n8: gb_pl:*\n9: gb_pr:*\n10: gb_ro:*\n11: gb_sts:*\n12: gb_sy:*\n13: gb_un:*\n14: gb_vl:*\n15: em_ba:*\n16: em_fun:*\n17: em_hum:*\n18: em_in:*\n19: em_mu:*\n20: em_om:*\n21: em_or:*\n22: em_ov:*\n23: em_pac:*\n24: em_ph:*\n25: em_pl:*\n26: em_ro:*\n27: em_sts:*\n28: em_un:*\n29: em_vl:*\n30: em_htg_hum:*\n31: em_htg_inv:*\n32: em_htg_other:*\n33: em_htgo_inv:*

BEST AVAILABLE COPY

C	1	129.4	6.0	10689	1	AE004008	AE004008 Xylella f
C	2	43.2	2.0	125020	3	AF429315	AF429315 Homo sapi
C	3	42.2	2.0	1461	3	AY051790	AY051790 Drosophila
C	4	42.2	2.0	76748	2	AC014256	AC014256 Drosophila
C	5	42.2	2.0	148847	3	AC008234	AC008234 Drosophila
C	6	42.2	2.0	162593	3	AC008356	AC008356 Drosophila
C	7	42.2	2.0	224400	3	AE003682	AE003682 Drosophila
C	8	41.6	1.9	720	8	CNS01D8A	AL116606 Botrytis
C	9	41.1	1.9	1700	1	RHMTNRC	M15810 R.melliloti
C	10	41.1	1.9	329100	1	SME591787	AL591787 Sinorhiz
C	11	40.4	1.9	125020	9	AF429315	AF429315 Homo sapi
C	12	40.0	1.9	4314	5	AF051784	AF051784 Xenopus l
C	13	39.4	1.8	3060	8	EMERLBA	L24395 Emerlicella
C	14	38.8	1.8	348450	1	MLBPRM4	AL583920 Mycobacte
C	15	38.6	1.8	10624	1	PD50X	X79242 Paracoccus
C	16	38.6	1.8	156971	2	AC099055	AC099055 Homo sapi
C	17	38.4	1.8	100306	2	AC105363	AC105363 Oryza sat
C	18	38.2	1.8	190050	2	AL646080	AL646080 Ralstonia
C	19	37.8	1.8	31896	2	AC017879	AC017879 Drosophila
C	20	37.8	1.8	172904	3	AC007414	AC007414 Drosophila
C	21	37.8	1.8	260910	3	AE003831	AE003831 Drosophila
C	22	37.4	1.7	38408	2	AC019786	AC019786 Drosophila
C	23	37.4	1.7	146153	3	AC007452	AC007452 Drosophila
C	24	37.4	1.7	203356	2	AL669952	AL669952 Mus muscu
C	25	37.4	1.7	264646	3	AE003818	AE003818 Drosophila
C	26	37.2	1.7	133691	9	AC074347	AC074347 Homo sapi
C	27	37.2	1.7	170398	2	AC013540	AC013540 Homo sapi
C	28	37.2	1.7	329709	1	AP002997	AP002997 Mesorhizo
C	29	36.8	1.7	810	6	AX110931	AX110931 Sequence
C	30	36.8	1.7	2016	8	PAN296102	AJ296102 Podospira
C	31	36.8	1.7	10925	1	AE004750	AE004750 Pseudomon
C	32	36.8	1.7	128135	2	AC104321	AC104321 Oryza sat
C	33	36.8	1.7	307150	1	CNSPAX01	AJ248283 Pyrococcu
C	34	36.8	1.7	349980	6	AX041106	AX041106 Sequence
C	35	36.6	1.7	1129	8	AF178952	AF178952 Zea mays
C	36	36.6	1.7	1496	8	CRENOLM	X66412 C.reinhardt
C	37	36.6	1.7	6709	14	MOCDRNAP	L38927 Molluscum c
C	38	36.6	1.7	170630	2	AC016905	AC016905 Homo sapi
C	39	36.6	1.7	185097	2	AC023148	AC023148 Homo sapi
C	40	36.6	1.7	190289	14	MC060315	U060315 Molluscum c
C	41	36.6	1.7	212956	2	AC090543	AC090543 Homo sapi
C	42	36.6	1.7	222759	2	AC073177	AC073177 Homo sapi
C	43	36.4	1.7	858	9	AF197129	AF197129 Homo sapi
C	44	36.4	1.7	1091	9	AY064416	AY064416 Homo sapi
C	45	36.4	1.7	1503	9	AK000576	AK000576 Homo sapi

ALIGNMENTS

RESULT 1\nAE004008/c\nLOCUS\nDEFINITION\nACCESSION\nVERSION\nKEYWORDS\nSOURCE\nORGANISM\n\nREFERENCE\nAUTHORS\n\n1 (bases 1 to 10689)\nSimpson,A.U., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M.,\nAlvarenga,R., Alves,L.M., Aaya,J.E., Bala,G.S., Baptista,C.S.,\nBarros,M.H., Bonaccorsi,E.D., Bordin,S., Boye,J.M., Britones,M.R.,\nBuono,M.R., Camargo,A.A., Camargo,L.E., Carro,D.M., Carrier,H.,\nColauto,N.B., Colombo,C., Costa,F.F., Costa,M.C., Costa-Neto,C.M.,\nContinho,L.B., Cristofani,M., Dias-Neto,E., Docena,C., El-Dorrry,H.,\nFacincani,A.P., Ferreira,A.J., Ferreira,V.C., Ferro,J.A.,\nFraser,J.S., Franca,S.C., Franco,M.C., Fritume,M., Furlan,L.R.,\nGarner,M., Goldman,G.H., Goldman,M.H., Gomes,S.L., Gruber,A.,\nHo,P.L., Hohnsels,J.D., Jungueira,M.L., Kemper,E.L., Kitajima,J.P.,\nand Marino,C.L.

TITLE The genome sequence of the plant pathogen *Xylella fastidiosa*. The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Sequencing and Analysis
 JOURNAL Nature 406 (6792), 151-157 (2000)
 MEDLINE 20365717
 PUBMED 10910347
 REFERENCE 2 (bases 1 to 10689)
 AUTHORS Simpson,A.J.G., Relnach,F.C., Arruda,P., Abreu,F.A., Acencio,M., Alvarenga,R., Alves,L.M.C., Araya,J.E., Bala,G.S., Baptista,C.S., Barros,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Brites,M.R.S., Bueno,M.R.P., Camargo,A.A., Camargo,L.E.A., Carraro,D.M., Carter,H., Colauto,N.B., Colombo,C., Costa,F.F., Costa,M.C.R., Costa-Neto,C.M., Coutinho,L.L., Cistofani,M., Dias-Neto,E., Docena,C., El-Dorri,H., Facincan,A.P., Ferreira,A.J.S., Ferreira,V.C.A., Ferro,J.A., Fraga,J.S., Franco,S.C., Franco,M.C., Frohme,M., Furlan,L.R., Garnier,M., Goldman,G.H., Goldman,M.H.S., Gomes,S.L., Gruber,A., Ho,P.L., Hohnselt,J.D., Junqueira,M.L., Kemper,E.L., Kitajima,J.P., Krieger,J.E., Kuramae,E.E., Laigret,F., Lambais,M.R., Leite,L.C.C., Lemos,E.G.M., Lemos,M.V.F., Lopes,S.A., Lopes,C.R., Machado,J.A., Machado,M.A., Madeira,A.M.B.N., Madeira,H.M.F., Marino,C.L., Marques,M.V., Martins,E.A.L., Martins,E.M.F., Matsukuma,A.Y., Menck,C.F.M., Miracca,E.C., Miyaki,C.Y., Monteiro-Vitorello,C.B., Moon,D.H., Nagai,M.A., Nascimento,A.L.T.O., Netto,L.E.S., Nhani Jr.,A., Nobrega,F.G., Nunes,L.R., Oliveira,M.A., de Oliveira,M.C., de Oliveira,R.C., Palmeri,D.A., Paris,A., Peixoto,B.R., Pereira,G.A.G., Pereira Jr.,H.A., Pesquero,J.B., Quaggio,B.R., Roberto,P.G., Rodrigues,V., de M. Rosa,A.J., de Rosa Jr.,V.E., de Sa,R.G., Santelli,R.V., Sawasaki,H.E., da Silva,A.C.R., da Silva,F.R., da Silva,A.M., Silva Jr.,W.A., da Silveira,J.F., Silvestri,M.L.Z., Siqueira,W.J., de Souza,A.A., de Souza,A.P., Terenzi,M.F., Truffi,D., Tsai,S.M., Tsubako,M.H., Vallada,H., Van Sluys,M.A., Verjovski-Almeida,S., Vettore,A.L., Zago,M.A., Zatz,M., Zedler,J. and Setubal,J.C.
 TITLE Direct Submission
 JOURNAL Submitted (02-JUN-2000) Organization for Nucleotide Sequencing and Analysis, Bioinformatics Lab - IC/Unicamp, C.P. 6176, Campinas, SP 13083-970, Brazil
 FEATURES
 source 1. 10689
 Location/Qualifiers
 /organism="Xylella fastidiosa 9a5c"
 /db_xref="taxon:160492"
 /clone="9a5c"
 complement(119. 367)
 /gene="XF1871"
 complement(119. 367)
 /gene="XF1871"
 /note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAF84677.1"
 /db_xref="GI:9106962"
 /translation="MTMTLKEDKTEIIDEQRIQAKLISEAKIQAEIHEVNAHTQKL MKETLKLTESKMYPVVGVSSLIAGATAATLFLFH"
 complement(430. 618)
 /gene="XF1872"
 complement(430. 618)
 /gene="XF1872"
 /note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAF84678.1"
 /db_xref="GI:9106963"
 /translation="MMEPARATAPOGGLASDPRTDAPKKTGLRKRKRYPTGSASIVLRHVLTORODKRRASP"
 941. 1288
 /gene="XF1873"
 941. 1288
 /gene="XF1873"
 /note="similar to GI1479340 (percent identity: 36 %/query

alignment coverage: 87.8 %/subject alignment coverage: 90.2 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AAF84679.1"
 /db_xref="GI:9106964"
 /translation="MSKNKINQIITGMAGEFQFNKLTVPERTIEQMLRLSLF CTIAREQKMEEPARATAPOGGLASDPQRTTDLQKTVSHKRRHAAETASIIQRLQ VLIQROKRRASP"
 1351. 1662
 /gene="XF1874"
 1351. 1662
 /gene="XF1874"
 /note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAF84680.1"
 /db_xref="GI:9106965"
 /translation="MTMTLKEDKTEIIDEQRIQAKLISEAKIQAEIHEVTHQEM NAHINEMARTQEMNARTQKLIETQKSKESQWYPVVASGLLAAGATAATLFLKLF H"
 complement(1839. 2918)
 /gene="XF1875"
 complement(1839. 2918)
 /gene="XF1875"
 /note="similar to SPIP03626 (percent identity: 36 %/query alignment coverage: 88.0 %/subject alignment coverage: 105.0 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"
 /codon_start=1
 /transl_table=11
 /product="phage-related protein"
 /protein_id="AAF84681.1"
 /db_xref="GI:9106966"
 /translation="MLNLVTGAPGNGKTLVAVDMLRIQIEIDKSLVKSAGVPRSFYTD IEGPVEAVRRLTGTVQVQASAPEDMETTPGQSVIYDEAHMFPAGRPGRSDPPYCDL DTHRRGCGDLAFYQWPTKIHHELRLNGEYHILNRANGLOTAGIYKRSRQDDPYDI HOREAEIEVWKFPKDRYALYASSLTLVSHKFRIPKRVMSALSCVYIGLVFHW YSPAHLEASASVAGASGASLRAAPSLSSSRSLVSGMRTYAVLETESAPTLSCGV SDRSCRCFTNDGYOIDMSVVECRILASPLPFVHYHAYVTSSSAAAVSSPSSPS SSVSISSSVSASAPLIGSSVPSH"
 complement(2918. 3202)
 /gene="XF1876"
 complement(2918. 3202)
 /gene="XF1876"
 /note="similar to SPIP03625 (percent identity: 36 %/query alignment coverage: 85.1 %/subject alignment coverage: 86.0 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"
 /codon_start=1
 /transl_table=11
 /product="phage-related protein"
 /protein_id="AAF84682.1"
 /db_xref="GI:9106967"
 /translation="WFSWMLMLRNIIQGTYARGLVAGLALVTTVPILPVTSALNL IVSKMSGISADVNLNLMGFGEALSIIGSAMLTRLAQSLHVGIVKATY"
 complement(3195. 4412)
 /gene="XF1877"
 complement(3195. 4412)
 /gene="XF1877"
 /note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAF84683.1"
 /db_xref="GI:9106968"
 /translation="WCSRSFVLYFTISALGLTLVTSFPFCLFAQNVPTVEVPTSVRY SSVISDGVSTATFEARSTALVNGVRYTVVDISASTIGSLAKVAVRGMAYNVYS


```

MSGLINGAWVIDELTHEVNSPALKEIPVGTAVAMMORPDGHNVEYSVTPQGLIAP
INAYOSLSPPOLRYLSSGSPSSVSESRNIVLHEGGGEYGHLSKTDQASVDPSSGILP
SVIIDPDQLQVLRSDPSRYNANVLTPDSORICVHLPEIYSALMKLRSLDEDEKASAP
DOOPSSGASPPSSSGTAMPSFCSMASVYCCFIDMWSDEFLKRPVLPVDPVYDKL
PQAKTWSSGLGEGACPSPTTPIEFSGYKTSVEISYOPCEDFAALMPVIVIAITILA
AVIAGCFRGKVNK"
/sequence="X1878"
complement(4481..4702)
/gene="X1878"
complement(4481..4702)
/note="hypothetical protein; identified by sequence
similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA084684.1"
/db_xref="GI:9106969"
/translation="MKSVIDRGIAFFSLSTSLVFAADNAGASEFDAGAAVTALGALS
GAVLIGAKLAPALISVGMKWLKGAIFG"
complement(4724..4924)
/gene="X1879"
complement(4724..4924)
/gene="X1879"
complement(4724..4924)
/note="hypothetical protein; identified by sequence
similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA084685.1"
/db_xref="GI:9106970"
/translation="MAOOLMSLYCTQYDVEARTCSQOAMVPPSLPLPSYEDVRILL
PILVMFLVAMEFHLFTVYVRD"
complement(4931..5236)
gene
Query Match 6.0%; Score 129.4; DB 1; Length 10689;
Best Local Similarity 44.9%; Pred. No. 1,1e-22;
Matches 888; Conservative 0; Mismatches 1051; Indels 38; Gaps 9;

```

```

722 AATACAGCAAGCAGCAATTAACCCCTATTAAGCCCGTTACTTCGTCGCGATCCATGCAAG 781
9787 GCTTCTCTTAAGAAACATACATCCCTACCGCTCCAAACAGCTGGCTGAATTTAGGATCAAC 9728
782 GCTACAAAGCGCTGACGACTATGACATGACATCCGTTTCCCGGGCAGTACGATCGCTAC 841
9727 CATTAGGTATGGTGTGACTTCGTGATGTGCTGCTGTTATCCAGAGACGTACCAATCGTTATG 9668
842 TCACCTCTTGGGGTGTGGAGAGATGATTCGAAACAGAAACAAATCCTGATCGAAGTTTC 901
9667 CTTTGTGGCCGCAATTTAGAAATAGTCATCAGACTTATCCGATGATGGCAGCATTC 9608
902 GCGGTATCAAGCAAGCATGTGGAAGAGCCATGAGCCGAGATCAGGCTATCCCGATCA 951
9607 TCAGAAATCTCATTTGCTTTGATTTAGGCGGCTGCAAAACGAATCCAGACATTCAGATGA 9548
962 AATATGCCACAGATATAGCTCAGAGTGCCTAATTTGGAAGAAATTCGATCGATGAAACC 1021
9547 AATATGGAGACCTTTATAGTGTGATTAATTAACATCGAAGAAACTTTGATGTCATTTGG 9488
1022 GCGGTCTCGCTGCTTGTGACGTATGATGATGATGATGATGATGATGATGATGATGATG 1081
9487 ATGGCTCCGACCCATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 9428
1082 ACTGATCCCTGATGAAGCGCAAGAGTGTGATGATGATGATGATGATGATGATGATGATG 1141
9427 CTTGTTGTAAGAAACAGAGGGGATACGTGACACAGAAAGATTTGGCTGCATCAAACTTGG 9368
1142 AGCCTTATTAAGGAAGAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCT 1201
9367 TTATCTGACAGAAACATACAAAGCAATACGATGATGATGATGATGATGATGATGATGATG 9308
1202 TCG---GTGTTACGAGGTGTTGCTTTTGCACAGTTTGCCAGCGATTGGCTACAAATC 1258
9307 ACCGAAAGTGGTGTCAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 9248
1259 CTGATGCTCATGCGCGATTCCTCAATTCGTTACAGCAAGTACAAAGATCTACCTCCCT 1318
9247 AATCTGATGCCCAACGTAAGCGGATATCAGAGCGCGATTCGCCCAACCATTTGAAGGCA 9188
1319 CGCTCGACCGCTAAGGTGCTCCCGCATGCTCATATTTGACCGCGCGATTCCTCCGCG 1378
9187 ACTTAAAGCAGATGAGCGACGTTACTTACCGGAATGATTCGCCAG---ATGACGCAAT 9131
1379 ACAAGCTCCCGATATATTCAGAAATGTATGACAA-----GAAATTCAAAGGC 1428
9130 ACTGCTCACCGAATACAAACAGCTTCCTCAACAGCGCTGAGCGCAATTTGATGTTT 9071
1429 GACACGAAAGATATGACAGCTTCGATTTGACAAAGGTGTTGCTTTATACGACAG 1488
9070 GGTGGAGATGGCATCCGCAACTCTCAAGGCTTGGGTGACACCCTTAAGTAAAGTAAAGTAA 9012
1489 TTCCATGCGCATGCTCAAGTGCATGACAGCAAGAAAGTTTCCAGAGCTATCGAAGAAAT 1548
9011 TCTGAGAAAGGCTGAAAGTGTATTAATGCCACCGTACAGCTTCGAAACAGTCAAGAT 8952
1549 CCGCAGATAGAGCTTTCCAGAGCGTAAATAGCTGTGCTCCGCTATTTCAAGCCGATGCG 1608
8951 CCAGCGATCGCTTATGCTGTGGCCATCATCAGCATTAATGAGATTAAGACGTAGAAAT 8892
1609 ATGGCCATGCTTATGCTTGAAGAGGCAAGCGTCTTTTCTTTGCGCGTTGGCGTGAAG 1668
8891 AAAATCCGTACTGAGAACTACTTAAGCCGTCGATCTGCAAGCTTCAAGCTTACGCTG 8832
1669 ATGTACCCCGGAGCTG-----CTGCGCAGAGATGCAACTTCAACATGCGATGAGC 1722
8831 TACAAACAGAGTATGCAAAATTTGTTTACCCAGAGCGCAACTCATCATATTAAGTATTAAC 8772
1723 TACGCTCATCAAGGATATGAAGCGCAGAGCGTGGCTGTGACAACTATCATATACGAC 1782
8771 TTGCGCACAGTGAAGGCTATTCACCCAAAGATGGCGTGAAGTACACACGTTTACGAC 8712

```

QY 1783 GGCAGGGCGATATGGAGACAGAGATCCCTTAAGACGATGAGTTTGCCGTACAGGAGAT 1842
 Db 8711 CTCGAGGGTGTGTAAGCGA---AGATACAGGGCGTGGAGCCATTGTGATTACCGAAGTCA 8655
 QY 1843 ATCCCTGACCTCTTCCGACACCAAAACATGATGGTCCCTATGCCA-----GACAGGTAG 1896
 Db 8654 CTGATTAAATGGCGATTAAGGCAAAAGAGCTACGCCAATTAGCCGATCAGCGATTGGCACT 8595
 QY 1897 CTCCATATCGCTTTCCTATCGAACAACAGCATCAGGGCGGTAACTCCGTTAGCCCCGTA 1956
 Db 8594 GTCCCTGTGAATTTTCTATCTGACATTGGACAATTACTGGCGGCACTCCGGATGGCGAGTC 8535
 QY 1957 TTTCATTAAGAAAGCCCGCTCTGATCGCTCTTGCTTTCGATGGCATCTGGAGACTTATGAGT 2016
 Db 8534 TTGATGTCGCCAGGAAAAATTGGTGGGTTTAGCATTTGACGGTAATTGGAGACTCTGCAGT 8475
 QY 2017 GGTGATCATCGATTCGAAACCCGATTCGACAGGCGCAAAATCAGGTTGGAGATCTCCGTCAGT 2076
 Db 8474 AGTAATTTGGGTTTGTGATCTCTGTGATCATCGAACAATTGCAAGTTGATAGTGTATATG 8415
 QY 2077 CTCCTCATGATTTGACAAATGGGGGTAGTGGCCCCGCTCATCTCAAGAGATGAAGTTG 2133
 Db 8414 CAGTGGATCATGACTGAAGATTGCAACACGACCCACACTTCTACTTAAGAAATTTGATTTTG 8358

RESULT					
2					
AF429315/c					
LOCUS	AF429315	125020 bp	DNA	linear	PRI 18-JAN-2002
DEFINITION	Homo sapiens junctionophilin 3 (JPH3)		gene,	partial cds.	
ACCESSION	AF429315				
VERSION	AF429315.1	GI:17646244			
KEYWORDS					
SOURCE	human.				

REFERENCE	Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Primates; Carnivora; Hominoidea; Homo.
TITLE	1 (bases 1 to 125020)
JOURNAL MEDLINE	Holmes,S.E., O'Hearen,E., Rosenblatt,C., Callahan,C., Hwang,H.S., Ingersoll-Ashworth,R.G., Flesher,A., Stevanin,G., Brice,A., Potter,N.T., Ross,C.A. and Margolis,R.L. A repeat expansion in the gene encoding junctionphillin-3 is associated with Huntington disease-like 2 Nat. Genet. 29 (4), 377-378 (2001) 21583737 11694876
PUBMED	2 (bases 1 to 125020)
REFERENCE	Holmes,S.E., Ingersoll-Ashworth,R.G., Ross,C.A. and Margolis,R.L. Direct Submission Submitted (03-OCT-2001) Psychiatry, Johns Hopkins Medical Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA Location/Qualifiers 1..125020
FEATURES	
SOURCE	

```

/ad_xiet="taxon:9606"
/chromosome="16"
/amp="16q24.3: between D16S520 and WI-12410"
/note="Isolated from a patient with Huntington's
disease-Like 2 (HDL2)"
complement(35581..35746)
/rpt_type=tandem
/rpt_unit=ctg
complement(<35507..>36887)
/gene="JPH3"
/product="junctophilin 3"
complement(<36507..>36887)
/gene="JPH3"
/note="JP3"
complement(<36507..36887)
/gene="JPH3"
/note="component of the junctional complex between plasma
membrane and endoplasmic reticulum"
/codon_start=1
/product="junctophilin 3"

```

BASE COUNT	ORIGIN
29056 a 32731 c 30696 g 28283 t 4254 others	<div> <div> /protein_id="AAI4094.1" /db_xref="GI:17646245" /translation="MSSGGRRFNPDDGGSYCGMEDGKAHGHGVCOTGPBGQGGEYTGSM HGFFLVGVYTMPSPGNYOQMGACCKRIGLIESGKKYVYGEWTHGKKKGYGVREACG NGAIEYESTMSGLDDGIGETIYDSG" </div> </div>

Query Match: 2.0%; Score 43.2; DB 9; Length 125020;
Best Local Similarity 9.8%; Pred. No. 5.2;
Matches 46; Conservative 226; Mismatches 194; Indels 4; Gaps 1;

250 ATCTTACCAACCACTGCGATACGGTCTATCCAGAGCCAAAGCAGGTGATCAC 309

Db 51261 AKCWMWYSYRKKCYCMSMNMASNAAMAMMMCTAAWYRKKTTRRRRSKSSSSYAYAMYAT 51202

[illegible][illegible]

51141 SBWTTSSMACVWYKGSBPKTKSMATSGCMBMGAMBRSGKGMRYWSCRYKGMRGBWGSgy 51082

430 AAGGATCACTGACGAGATGGAGCGTCTGCCCAAAGCTCAGGAGTATGCCAAGACTG 489

Db 51081 KGCSGYMSMGMYGSGRTSKYCSYTGGYCSKCSAKMCKSKSTSKSCCTKSKYS---CSMG 51026

490 GCCAAAGAATGCAGACGAGAACCACTCTGCATCGTAGAGCCTTCTATTCCAAC 549

Db 51025 RYSSYCTCWGGRMCKGKSSCCASRSRGCKSYKMAGSWRCRGRGMSWCYWSSMKSYSMAS 50966

350 AACGAAIACITCCICAIAGICACGAIGAIICAAAGACGIIAGIAGGAIIGCICCI 009

[illegible]

```
DB      50905 MMBDMYYDKMBAKVSSGTPCYHHBSAWMCKSMSBPSDSCGYSGMYHMKSYWKKSSASKHMY 50846
```

670 GGGACTTCAGCGTATCCGGGTATGCGCGTGCCGACACCGGCGGC 719

Db 50845 SGMKSSVYBMDCMSMYTSBSKSVRMBRRWGSYBRYRKCAYMRYBHRSS 50796

U
C
C
T
T
E
J

AY051790	LOCUS	AY051790	1461 bp	mRNA	linear	INV 27-AUG-2001

DEFINITION	Drosophila melanogaster LD30634 full length cDNA.
ACCESSION	AY051790

VERSION 01.10224000
KEYWORDS FLI_CDNA.

ORGANISM Drosophila melanogaster
Eukaryota: Metazoa: Arthropoda: Tracheata: Hexanoda: Insecta:

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE
AUTHORS
1 (pages 1 to 1401)
Stapleton, M., Brokstein, P., Hong, L., Agbayan, A., Carlson, J.,

Gonzalez, M., Guarín, H., Li, P., Liao, G., Miranda, A., Mungall, C.J.,

TITLE
Direct Submission
Yu, C., Lewis, S.E., Rubin, G.M. and Celniker, S.

JOURNAL
Submitted (10-AUG-2001) Berkeley Drosophila Genome Project,
Lawrence Berkeley National Laboratory, One Cyclotron Road,

COMMENT
Sequence submitted by: bernerey, ca 94120, usa

Lawrence Berkeley National Laboratory
Berkeley, CA 94720
This clone was sequenced as part of a high-throughput process to
sequence clones from Drosophila Gene Collection 1 (Rubin et al.,

DEFINITION	Drosophila melanogaster genomic scaffold 142000013386035 section 7
ACCESSION	AE003682 complete sequence.
VERSION	AE003682.2
KEYWORDS	GI:10726402
SOURCE	HTG.
ORGANISM	fruit fly. Drosophila melanogaster Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE	1 (bases 1 to 224400)
AUTHORS	Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F., George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N., Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X., Brandon,R.C., Rogers,Y.H., Blazer,R.G., Champe,M., Pfeiffer,B.D., Man,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor Miklos,G.L., April,J.F., Agbayani,A., An,H.J., Anderson,P., Baxendale,C., Baldwin,D., Ballew,R.M., Basu,A., Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y., Benos,P.V., Bereman,B.P., Bhendari,D., Bolshakov,S., Borkov,D., Botchan,M.R., Bouch,J., Brockstein,P., Brothier,P., Burris,K.C., Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I., Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de Pablo,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M., Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C., Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferreira,S., Fleischmann,M., Foster,C., Gabriellian,A.E., Gard,N.S., Gelbart,W.M., Glasser,K., Glodok,A., Gong,F., Gorrell,J.H., Gu,Z., Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J., Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J., Wei,M.H., Ibegwam,C., Jalali,M., Kalush,B.E., Kodira,C.D., Kraft,C., Kraevitz,S., Kulp,D., Lai,Z., Laake,P., Lei,Y., Levitsky,A.A., Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C., McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C., Morris,J., Mosher,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L., Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K., Nusker,D.R., Pacle,J.M., Palazolo,M., Peltman,G.S., Pan,S., Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K., Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I., Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C., Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R., Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A., Weinstock,G.M., Weissenbach,J., Williams,S.M., Woodage,T., Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F., Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H., Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O., Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C. The genome sequence of Drosophila melanogaster Science 287 (5461), 2185-2195 (2000) 20196006 2 (bases 1 to 224400) JOURNAL MEDLINE AUTHORS TITLE JOURNAL COMMENT FEATURES SOURCE
gene	location/Qualifiers 1..224400 /organism="Drosophila melanogaster" /db_xref="taxon:7227" /chromosome="3R" complement(join(235..1324,1643..2221,2355..2417)) /gene="D1" /product="Ct27541" /db_xref="FLYBASE:FBan0009745" /db_xref="FLYBASE:FBgn0000412" complement(<235..>2417) /gene="D1" /note="CG9745" /map="85D2-85D3" /db_xref="FLYBASE:FBan0009745"
gene	complement(join(672..1324,1643..2057)) /gene="D1" /note="D1 gene product" /codon_start=1 /db_xref="FLYBASE:FBan0009745" /db_xref="FLYBASE:FBgn0000412" /protein_id="AAF54341.1" /db_xref="GI:7299143" /translation="MEEVKKRRGRPKASVGSSTAAVAALISPGIKRGRPAKNG SSGGGGRGRPRPKAKIONDEPEDEGGDGGSGAELANSSPSPPKGRKKS SGAGSSGSVYTPGSAKRRKRRKHDPSSNEDDDDDDDDNSSISERRPGR SGSVNLNTSRGGRGRKRRKRVNSGGEEVQVRKRRPPQNSGGSGSTGYPTG RRGRKANAAVEERKEDDDDDDDDDSSGEESSEPKTVAPKRRGRSLAAGVSK EETTPRRSPRPAKNIIDDADSDAOGGNSKESNEEDRAVDGTPTKGDKLKNVSDG ENDANGYVSDNYNSESVA" complement(join(3004..3235,3300..3914,3972..4605, 4865..5199,5255..6265,6330..7367,7561..7747,7809..>7907))) /gene="CG9746" /product="Ct27547" /db_xref="FLYBASE:FBan0009746" /db_xref="FLYBASE:FBgn0037663" complement(<3004..>7907) /gene="CG9746" /map="85D2-85D4" /db_xref="FLYBASE:FBan0009746" /db_xref="FLYBASE:FBgn0037663" complement(join(3004..3235,3300..3914,3972..4605, 4865..5199,5255..6265,6330..7367,7561..7747,7809..7906)) /gene="CG9746" /note="CG9746 gene product" /codon_start=1 /db_xref="FLYBASE:FBan0009746" /db_xref="FLYBASE:FBgn0037663" /protein_id="AAF54342.1" /db_xref="GI:7299144" /translation="MGNQVJGAPSOIYAVEHYFSGQFGEIATSSNNGSTREFYVAK ATDEQIIVKFKVHKDPIPLEDHEKEREGIKTSLANAVCPQFQVELIDKAY IKREYKHLIDRVSRPLVYIEKKWIFOLICLNCHOKIGCHQIKLENIITS NMWIIIDASRKPPLYLPNDNDADYFEDTSRRRCVYAPRFPYKSLADDDGNGN NMYIHTDSIIRLAPSNCTLPAMDIFPSGACGALLEMTBGRAPPELSOLLARGER DIVERKHLAIEENRKLNLASIDHISNRKSAEYLDDERGQLEPEFYSLQSTILO MSSPPIIMPDPKIOCLNHDICIKVLVQADLPPEEKREKDPYLCALPEHD GILLITVYSCIRGLKQISKTIRIALELOKSKYTSSTIIRLPIYLHLAQSPA KVOYOAIEFLTCAGVCKVIPSIDANVPEYIIPSTIDASTSVIYVAVARATIAS LKSNAVYFLFEORNPANMPPRYEAELNADHIVROAVISLIDSDOYVQNTMES GICDCAFPGRKANDYLSHMTFLNDEKDKLRGAFYDNTINGVAGYQKMSDITLP LIQGGTIDREERVAIAKAVTILIEGRIKPAVETIYQETACIHCHEKLNWREHC GIATTARNLASIDVQCKTMAPIGAFKILQVERPHILDCVHPVPAPROIFDSLR FDDIHFIETALRARSRYRSQTRQALPOYEENGQTLNFRRLSEGLDILEMOLA NMPFLISMKHALODIDDPASNGRIVRSQVACHPEYIADAKASKMPLDNSSGPT LASPASDAITPFLGAGGAGSPASGAVGVPTVYSAATSLGETYMPRNCWLEERS EKRKDEESTAKIKSRVYNVLASQCSYKVTYPPRPGRMGTIVLHHEHSVVK LMSLRPHGSLFASGIDGTVRLMDSKLNQNGVNSKRVYSAKPIYLAHLSADGQS LAVGKDGSMAMRILDRNSAKMTLQALQMNQNDKSGPVVDMHAYQAOQSVYVATLY GGVAVMDTQMSAMRILQNELHGHVYTTICAPDTGSMLATGSGGKILCWDIREFPLI AELIRPADSWIKRVACHPEPSYILASQSNNEVSYNLETGORTVYLAASPVALS ASPSDPAITGILISGVNDSPEPLINGSSORTIRYMDITTPKNSILKVPANNDLADV FNYGARLIDGSOYIEQIITISMASTGDSQRLRSSTENPNSGDPMTAFASHHDIITPLM CKDKQIYIYASASRNVIKLMK" join(9619..10588,10978..11126,12073..12352,12809..13132, 13366..13591) /gene="CG8420" /product="Ct24697" /db_xref="FLYBASE:FBan0008420" /db_xref="FLYBASE:FBgn0037664" /db_xref="FLYBASE:FBgn0037664" /gene="CG8420" /map="85D2-85D4" /db_xref="FLYBASE:FBan0008420" /db_xref="FLYBASE:FBgn0037664" join(9827..10588,10978..11126,12073..12352,12809..13132, 13366..13446)
CDS	
gene	
mRNA	
CDS	

JOURNAL	J. Bacteriol.	169,	1422-1432	(1987)
MEASURE	871657/45			
FEATURES				
source	location/Qualifiers			
	1. .1700			
	/organism="Sinorhizobium meliloti"			
	/db_xref="taxon:382"			
CDS	<1. .255			
	/note="ntrb"			
	/codon_start=1			
	/transl_table=1			
	/protein_id="AAA26345.1"			
	/db_xref="GI:152392"			
	/translation="EFCVHDNGEVPDPDLPHFDPIITTTKNGSGGLALVAKIIGG			
	HGGIVECDQSHSRITFRVLPMPASKGLADETPMTKTNG"			
	252. .1700			
CDS	/note="ntrc"			
	/codon_start=1			
	/transl_table=1			
	/protein_id="AAA26346.1"			
	/db_xref="GI:152393"			
	/translation="MTGATLLVADDDAIRTVLNOLSRAGYDVYRITSNATLWRWIA			
	AGDGLVTVTVVMPDENAFDLPRIKARPDLPVLVMSAQMTFMTAIKASEKGYDVIV			
	PKPDLDELIGITGRALAEPRKRPSSDSDSGMPVGRSAAMQELIVRYARLMTQND			
	LPTMTGSESGTGEKIVRALHDYGRKRNQSPVNAIMNAIPIRDLISELFGREKATFG			
	AQRSTRERFQAEBSGTLFDEIGDMPDAQOTRLRLVLAQGGFTVVGRTPIRSQVRI			
	AATKDLKQSIINOGLFREDYLRNLVPLRLPDRADEDIPLDVRHFVQAEEGIDL			
	VKRPDQALELMKMPGWVRELENTVRLTALYPODVITREIENELSEIPDSPT			
	EKAARSGSLISQVAEENKROYFASGDALPESGLYDRVLAEMEPPLIALATATATRG			
	NOTFAADLLGINTRTLEKRIIREIGSVYRSIA"			
BASE COUNT	341 a	553 c	511 g	295 t
ORIGIN				
Query Match	1.9%;	Score 41;	DB 1;	Length 1700;
Best Local Similarity	55.2%;	Pred. No. 14;		
Matches	80;	Conservative	0;	Mismatches 65; Indels 0; Gaps 0;
OY	1654	GCCGCTTGCGTGAGATGTACCCGGGACGCTGCTGCGGAGGATGCCAACTTCACCATG	1713	
Db	681	GCCGCCACGACGAAATCTACCGCTGCTGCGCGCTGATGACACCGACCTCAGCGTG	740	
OY	1714	GGTATGACCTACGCGTCCATCAAGGGATATGAACCGCAGGACGGTGGTACAACTAT	1773	
Db	741	ATGATCACCGGGGAGCTCCGGCACGGCAAGGAACCTCGTTGCCCGTGCACCTGACGACTAT	800	
OY	1774	CATACGACAGCGCAAGGGCGGTATTGG	1798	
Db	801	GGAACGCGGAGAAATGGCCCTTCG	825	
RESULT 10				
SME591787				
LOCUS	SME591787	329100 bp	DNA	linear BCT 16-AUG-2001
DEFINITION	Sinorhizobium meliloti 1021 complete chromosome; segment 6/12.			
ACCESSION	AL591787			
VERSION	AL591787.1			
KEYWORDS	GI:15074266			
ORGANISM	Sinorhizobium meliloti.			
	Sinorhizobium meliloti.			
	Bacteria: Proteobacteria; alpha subdivision; Rhizobiaceae group.			
	Rhizobiaceae: Sinorhizobiaceae.			
REFERENCE	1 (bases 1 to 329100)			
AUTHORS	Capela,D., Barilley-Hubler,F., Gouzy,J., Bothe,G., Ampe,F., Batut,J.,...			
	Boistard,P., Becker,A., Bontuy,M., Cadieu,E., Dreano,S., Gloux,S.,			
	Godrie,T., Goffeau,A., Kahn,D., Kiss,E., Lelaire,Y., Maury,D.,			
	Pohl,T., Portetelie,D., Puehler,A., Purnelle,B., Rameger,U.,			
	Renard,C., Thebault,P., Vandenbol,M., Weidner,S. and Galibert,F.			
TITLE	From the Cover: Analysis of the chromosome sequence of the legume			
JOURNAL	Symbiot Sinorhizobium meliloti strain 1021			
	Proceedings of the National Academy of Sciences of the United			
	States of America. 98 (17), 9877-9882 (2001)			
PUBMED	11481430			
REFERENCE	2 (bases 1 to 329100)			
AUTHORS	Gouzy,J.			

TITLE	Direct Submission
JOURNAL	Submitted (26-JUL-2001) Gouzy J., Submitted on behalf of the MELLIO
COMMENT	EU Consortium MELLIO EU Consortium: Laboratoire de Biologie Moléculaire des Relations Plantes-Microorganismes, UMR215-CNRS-INRA, BP27, F-31326 Castanet, France, Laboratoire de Génétique et Développement UMR6061-CNRS, Faculté de Médecine, 2 avenue du Pr. Léon Bernard, F-35043 Rennes, France, GATC GmbH, Fritz-Arnold-Str. 23, D-78467 Konstanz, Germany, Universitaet Bielefeld, Biologie IV (Genetik) Universitaetsstr 25, D-33615 Bielefeld, Germany, unite de Biochimie physiologique, Universite catholique de Louvain, Place Croix du Sud 2, Bte 20, B-1348 Louvain-la-Neuve, Belgium, Unite de Microbiologie, Faculte des Sciences Agronomiques de Gembloux, Avenue Marechal Juin 6, B-5030 Gembloux, Belgium. E-mail:Jerome.Gouzy@toulouse.inra.fr http://sequenec.toulouse.inra.fr/melliot1.html.
FEATURES	Location/Qualifiers
source	1..329100 /organism="Storhizobium melliot1" /strain="1021" /db_xref="taxon:382"
tRNA	152..227 /gene="tRNA-THR_TGT" /note="codon recognized: UGU; predicted by tRNAscan-SE" /product="tRNA-Thr" /evidence="not_experimental 152..227 /gene="tRNA-THR_TGT" complement(373..798) /gene="SMC04435" complement(373..798) /gene="SMC04435" complement(373..798) /function="miscellaneous; hypothetical/global homology" /note="Product confidence : hypothetical Gene name confidence : hypothetical predicted by Homology" /codon_start=1 /transl_table=1 /evidence="not_experimental /product="CONSERVED HYPOTHETICAL PROTEIN" /protein_id="CAC45913.1" /db_xref="GI:15074267"
gene	/translation="MNNALVPEFAVSDMQSKPEYCNILGFSQDYPEREGCYSLSD GAEIMDIGRTGFADGHLPRSPFGRLNVCIRVASADALVRALDRGIALYLPLE ERWYRGDEEYGNRFVADPDGYLLRFREPLGRPRTS" complement(914..1855) /gene="SMC01330" complement(914..1855) /gene="SMC01330" complement(914..1855) /function="miscellaneous; not classified regulator" /note="Product confidence : putative Gene name confidence : hypothetical predicted by Homology predicted by Framed" /codon_start=1 /transl_table=1 /evidence="not_experimental /product="PUTATIVE TRANSCRIPTION REGULATOR PROTEIN" /protein_id="CAC45914.1" /db_xref="GI:15074268"
CDS	/translation="MODLIDIALFAAAGVVKHNGFSAARDLNVPSKLSKHVARLEEDD GVRLERSTRKLMTVEGRIFYEAGGVGVAAGAAEYAVRAEPTGVACPIGEF TPMLADLPAHRYRPGVRLITLITASNRIDLIERIDVALRAPOLDPSOLYKRG EVRRLAASPTLLRLRGEITTDNLSDPRTLSMNGCHNDYRLVHAGAEIARHY VGSCDFILERAAIEGMIALPDHICIERAETGALVPLPEWTSNVNVAHLVFPSSR GLPATRALIDFLAENLIKALERCREVDPRAAFET" 2063..2683 /gene="acpD OR SMC01329" 2063..2683 /gene="acpD OR SMC01329" /function="small molecule metabolism; fatty acid biosynthesis"
gene	
CDS	


```

gene      /strain="TN"
          /db_xref="taxon:1769"
          551..1786
          /gene="ML0842"
          551..1786
          /gene="ML0842"
          /note="Similar to Mycobacterium tuberculosis hypothetical
          44.6 kDa protein RV1464 or MT0007.11 TR:053155
          (EMBL:AL021184) (417 aa) fasta scores: E(): 0, 43.6% id in
          408 aa and to many other bacterial nifs-homologues, e.g.
          Bacillus subtilis YurW protein YurW TR:032164
          (EMBL:Z59120) (406 aa) fasta scores: E(): 0, 46.7% id in
          405 aa. Previously sequenced as TR:032975 (EMBL:Z36741).
          Contains Pfam match to entry PF00266 aminotran_5,
          Aminotransferases class-V.
          Similar to ML0117, ML0596 and ML1708"
          /codon_start=1
          /transl_table=11
          /product="conserved hypothetical protein"
          /protein_id="CAC31223.1"
          /db_xref="gi:13092923"
          /translation="MVDVEDIOAVRADPILQETVNGKPLIMPNMAPTQKPOVIDR
          LSYFAHNSNHRRAHRLARADAYEAETRRRTGAKAKEIIFVRETPIANIL
          VAYANGHLNPGDESVITLHEHNAIVPQOLSSQTEALIKAPVDDAGMLSEEE
          DLLGRTEKLVATQVSNAIGVTQGEKVELGHHYGRARVLIDGQAS.IPH.IINVELIG
          ADPEFSSEKIIYGPPIGIVLCEDEVLTEMPQGGMMIVDYVLESLSKVGPNKKEE
          AGTGNIAAVNGIGALRYVERGVORISHQOALDATTPLADIPGVRLVGTETKEEA
          ILSVLVINGHEPLEYKGLNAGLIVNRGHHCAQPLVRLRLGLEATVPSPFAFYTYEE
          IDVFIVNVRIRAEAGTGNG"
          653..1138
          /gene="ML0842"
          /note="Pfam match to entry PF00266 aminotran_5,
          Aminotransferases class-V, score 23.70, E-value 1.7e-08"
          1908..2395
          /gene="lppp"
          /note="ML0843"
          1908..2395
          /gene="ML0843"
          /note="Possible pseudogene of M. tuberculosis orthologue
          lppp (best blastx score 294)"
          /codon_start=1
          /pseudo
          /transl_table=11
          /product="lppoprotein (pseudogene)"
          2549..2555
          /note="possible RBS"
          2559..4112
          /gene="nark"
          /note="ML0844"
          2559..4112
          /gene="nark"
          /gene="nark"
          /note="Similar to Mycobacterium tuberculosis putative
          nitrite extrusion protein nark1 or RV2329c or MTG3612.05
          TR:P11883 (EMBL:Z79702) (515 aa) fasta scores: E(): 0,
          69.3% id in 488 aa and to Escherichia coli nitrite
          extrusion protein 2 narX SW:NARU.ECOLI (P37758: p77696)
          (462 aa) fasta scores: E(): 0, 36.6% id in 459 aa. There
          is a frameshift near the C-terminus relative to the M.
          tuberculosis homologue. Previously sequenced as TR:032974
          (EMBL:Z38741). Contains hydrophobic, probable
          membrane-spanning regions."
          /codon_start=1
          /transl_table=11
          /product="putative nitrite extrusion protein"
          /protein_id="CAC31225.1"
          /db_xref="gi:13092924"
          /translation="MDQLVLLQAEBSLYGPRRSKTPVIDRHSLKSSFFLRRLRAC
          RISHSDPEDQAAEAGNKTIARLNILNMLNGLVYHGYSWTLLPMELEMPDYGFSFA
          GDKPLATTATITVGCINRVPISIAATLEGGNMAIEGWTMLIPTIRNVIIAHPIGL
          IMPYLACALITGICGGNRAASMTNANATYPRKRLAGAGLGGAGNIGSVIYOVGML
          VIASVGDKRPITVCGIYVLLIINGIGMLPMNDIEHRIQVNTIRVLYFVVSRTOS
          UNVALLIYLAGSGSFIGFSFAGQVLETFVAGGOSTQOALAHAEIAFGTTLAAVAR
          EWGGLIARLRGSGRTLVVFGAMVFAAGLIGVLGIIIEGSRVCPIRGVMASYEGFTT

```

```

gene
GID1LRRESYLWTGGVTAFAEWFEMLCYAAAGVLTWKMTVCRPLDGNLHDEANAFAS
VGA$RTHRC"
/complement(4100. .4681)
/gene="ML0845"
/complement(4100. .4681)
/gene="ML0845"
/notes="unknown function. Similar to part of some acyl-CoA
oxidases e.g. Arabidopsis thaliana acyl-CoA oxidase cx2
TR:065201 (EMBL:AF057043) (692 aa) fasta scores: E():
1.5e-06, 35.2% id in 125 aa. Previously sequenced as
TR:032973 (EMBL:Z98741) (193 aa) fasta scores: E(): 0,
99.5% id in 193 aa."
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAC31226.1"
/db_xref="GI:13092925"
/translation="MPLRPTNAASERYTKRATETIIQRMVGARODESKGSLFNKGT
OCTIEKEDCDLISAVARRQVSKEMAFPTFMYVDHLHMANETDRVLTFLPLA
GIESNDPARKLIDVCDLVALSVKARGHSSIDTSPSLKAIAGSRMP$AALL
CNDVARRSESQSCVTPRCRSTRICPELDR"
/complement(4959. .5480)
/gene="ML0846"
/complement(4959. .5480)
/notes="Possible pseudogene similar to M. tuberculosis
paralogue RV1747 (Best blastx score 127)"
/codon_start=1
/pseudo
/transl_table=11
/product="ABC transporter (pseudogene)"
/complement(5522. .5848)
/gene="ML0847"
/complement(5522. .5848)
/gene="ML0847"
/notes="Possible pseudogene of M. tuberculosis orthologue
RV2327 (Best blastx score 184)"
/codon_start=1
/pseudo
/transl_table=11
/product="conserved hypothetical protein (pseudogene)"
/5906. .8080
/gene="ML0848"
/5906. .8080
/gene="ML0848"
/notes="Similar to Mycobacterium tuberculosis hypothetical
ABC transporter ATP-binding protein RV2326c or MTCY3612.08
SW:YV26.MYCrv (P71886) (697 aa) fasta scores: E(): 0,
76.9% id in 697 aa. Shares similar domains with many
ABC-type transporters e.g. Streptomyces roseofulvus Arpase
component of putative ABC transporter fmd TR:068910
(EMBL:AF058302) (524 aa) fasta scores: E(): 8.4e-14, 31.5%
id in 495 aa and Synecococcus sp. nitrate transport
APP-binding protein Nrd nrd SW:NRPD_SYNP7 (P38046) (274
aa) fasta scores: E(): 1.6e-12, 34.8% id in 198 aa.
Previously sequenced as TR:032971 (EMBL:Z98741). Contains
hydrophobic, possible membrane-spanning regions. Contains
2 pfm matches to entry PF00005 ABC_tran, ABC transporter.
Contains 2 x PS00017 ATP/GTP-binding site motif A
(P-loop). Contains 2 x PS00211 ABC transporters family
signature."
/codon_start=1
/transl_table=11
/product="ABC transporter"
/protein_id="CAC31229.1"
/db_xref="GI:13092926"
/translation="MDLCIHRMALLCOATRRCHRWKTKOLGRCAAPLADPTGIGVY
TAHDPRHILHRRSGLOBELAQASYLAKCAVTAIVSYVPEFAGLLITGVPVGLIA
YARRRVLTAAVAGVIAFLITIGLGGFTIYVANSYVIGGLGVYKRGQGTLYTALA
LEFAGLFAAGANVALVAVLGRHLIFKAMTANVDGIAATLTMHLPVAVOLKRYEAD
GIQHPMDKLGIVFTIILVLSIGMWYLSVLEIRIDIPDYKHTDASACNEDAPVP
VYVLDMPKVAEIRP$HAGODALREVSLDRLVGVHVAVTGANGSGKTTLMILILAGREPTSG

```

	misc_feature	/gene="Ml0848"	/note="Pfam match to entry PF00005 ABC_tran, ABC transporter, score 112.30, E-value 9e-30"	
	misc_feature	/gene="ML0848"	/note="PS00017 ATP/GMP-binding site motif A (P-loop)"	
	misc_feature	7109..7153		
Query Match	Best Local Similarity	1.8%	Score 38.8; DB 1;	Length 348450;
Matches 115;	Conservative	0;	Mismatches 127;	Indels 0;
Gaps		0;		
OY 1374	TGCGCACAAGCTCCCGCATATTCACAAATGTATTCGACCAAGAATTCAAAGGGACMC	1433		
Db 292422	TCCGCCAGTCTGATGCATCACACCCTTTTGTTGCGAACGSGCAGCTTCGCCAACA	292481		
OY 1434	GAAGCAATGATCGACAGCTCGATTTCGACAGAGCTGTGCTCTTATTAAGCAAAATTCCA	1493		
Db 292482	GCGCAAGTGGCTGACTTCTTTCNAAGTTACAAAGTTAGAACCTTGCATCGCGGTGTGA	292541		
OY 1494	TGCCATGCTCAAGTCCATGACACAGAAAAGTTGCCAAGGCTATTCGACAAAGATCCGC	1553		
Db 292542	TTTCGATTTGTGATGCCCCCGATGACCCCAATGGTGTGCTGCGGTGTCGCCAGCATCGGC	292601		
OY 1554	AGTAGAGCTTTCACAGACGGTAATAGCTGCTGCTCGCCTATTCAGCGCGATGCCATGGC	1613		
Db 292602	GGGATGGTGTGCTGCTGATCAATCCAGCATCCGATATTGATGCTGTGGCGACGC	292661		
OY 1614	CA 1615			
Db 292662	CA 292663			
RESULT 15	PDSOX/c	10624 bp	DNA	linear BCT 19-DEC-2001
LOCUS	DEFINITION	PDSOX		
VERSION	ACCSSION	X79242		
KEYWORDS	KEYWORDS	X79242.4 GI:18308976		
SOURCE	ORGANISM	Bacteroides dentrificans ORF1, ORF2, ORF3 and soxX, soxY, soxZ, soxA, soxB, soxC, soxD, soxE, (partial) soxF genes.		
REFERENCE	AUTHORS	J. Bacteriol. 176 (20), 6188-6191 (1994)		
REFERENCE	TITLE	Identification and sequence analysis of the soxB gene essential for sulfur oxidation of Paracoccus denitrificans GB17		
JOURNAL	REFERENCE	J. Bacteriol. 176 (20), 6188-6191 (1994)		
MEDLINE	AUTHORS	Wodara,C., Bardischewsky,F. and Friedrich,C.G.		
TITLE		Cloning and characterization of sulfite dehydrogenase, two c-type cytochromes, and a flavoprotein of Paracoccus denitrificans GB17: essential role of sulfite dehydrogenase in lithotrophic sulfur oxidation		
JOURNAL		J. Bacteriol. 179 (16), 5014-5023 (1997)		
MEDLINE		97405897		
REFERENCE		3 (sites)		
AUTHORS		Modara,C.		
TITLE		Direct Submission		
JOURNAL		Submitted (13-MAY-1994) C. Wodara, Institut fuer Technische		

REMARK	REFERENCE	AUTHORS	TITLE	JOURNAL
REMARK	REFERENCE	AUTHORS	TITLE	JOURNAL
4 (sites)	4 (sites)	Modara,C.	Direct Submission	Submitted (17-SEP-1996) C. Modara, Institut fuer Technische Mikrobiologie, Universitaet Dortmund, Fachbereich Chemiechnik, Lehrstuhl fuer Technische Mikrobiologie, Em11-F19ge-Str. 66, 44221 Dortmund, FRG
5 (sites)	5 (sites)	Friedrich,C.	Direct Submission	Submitted (25-JUN-1997) Friedrich C., Institut fuer Technische Mikrobiologie, Universitaet Dortmund, Fachbereich Chemiechnik, Lehrstuhl fuer Technische Mikrobiologie, Em11-F19ge-Str. 66, 44221 Dortmund, FRG
6 (sites)	6 (sites)	Friedrich,C.	Direct Submission	Submitted (10-JUL-1997) Friedrich C., Institut fuer Technische Mikrobiologie, Universitaet Dortmund, Fachbereich Chemiechnik, Lehrstuhl fuer Technische Mikrobiologie, Em11-F19ge-Str. 66, 44221 Dortmund, FRG
7 (bases 1 to 10624)	7 (bases 1 to 10624)	Friedrich,C.	Direct Submission	Submitted (09-JUN-2000) Friedrich C., Institut fuer Technische Mikrobiologie, Universitaet Dortmund, Fachbereich Chemiechnik, Lehrstuhl fuer Technische Mikrobiologie, Em11-F19ge-Str. 66, 44221 Dortmund, FRG
On Jan 24, 2002 this sequence version replaced gi:8517639.	Location/Qualifiers			
1. 10624	1. 10624			
/organism="Paracoccus denitrificans"	/organism="Paracoccus denitrificans"			
/strain="GB17"	/strain="GB17"			
/db_xref="taxon:266"	/db_xref="taxon:266"			
complement(320..769)	complement(320..769)			
/note="ORF1"	/note="ORF1"			
/codon_start=1	/codon_start=1			
/transl_table=1	/transl_table=1			
/product="hypothetical protein"	/product="hypothetical protein"			
/protein_id="CAB94376.1"	/protein_id="CAB94376.1"			
/db_xref="GI:8517640"	/db_xref="GI:8517640"			
/db_xref="SPTREMBL:O9LCV3"	/db_xref="SPTREMBL:O9LCV3"			
/translation="MTAPWIRDKIASOGSSGMCGRAPVPAFRKNGDPELGENTIR ODEPQEDFDALIKANDASTFEKALDHDRIMLTICYLMSGPPSYTELENLSSRAV VSOQLARHREGIVSARRDQITIFYSLIDRPVVDLLAVLKRLETPDC"	/translation="MTAPWIRDKIASOGSSGMCGRAPVPAFRKNGDPELGENTIR ODEPQEDFDALIKANDASTFEKALDHDRIMLTICYLMSGPPSYTELENLSSRAV VSOQLARHREGIVSARRDQITIFYSLIDRPVVDLLAVLKRLETPDC"			
1228..1965	1228..1965			
/note="ORF2"	/note="ORF2"			
/codon_start=1	/codon_start=1			
/transl_table=1	/transl_table=1			
/product="hypothetical protein"	/product="hypothetical protein"			
/protein_id="CAB94378.1"	/protein_id="CAB94378.1"			
/db_xref="GI:8517641"	/db_xref="GI:8517641"			
/db_xref="SPTREMBL:O9LCV2"	/db_xref="SPTREMBL:O9LCV2"			
/translation="MIDVYGGAAIGLISFLSPCLIPMYFYLISVYAGASVRELRDGGTVYAAAGRQRMVLRALAFALGVTTIFILMLDLGATLALGDAFRQKGTLSVRAAGVLYLFGGLHFLVLRIGLVLREARVQTEAKPSTLILGALVYGLAFGFGWPCVGPALAILVWASGMELMVEGGLLVYGLAMTLPVLAFAFPAPFLSWARRAVWGHEKAMGVALIVEFGILATISVNAVIDMMIRNEDMTATIL"	/translation="MIDVYGGAAIGLISFLSPCLIPMYFYLISVYAGASVRELRDGGTVYAAAGRQRMVLRALAFALGVTTIFILMLDLGATLALGDAFRQKGTLSVRAAGVLYLFGGLHFLVLRIGLVLREARVQTEAKPSTLILGALVYGLAFGFGWPCVGPALAILVWASGMELMVEGGLLVYGLAMTLPVLAFAFPAPFLSWARRAVWGHEKAMGVALIVEFGILATISVNAVIDMMIRNEDMTATIL"			
1979..2539	1979..2539			
/note="ORF3"	/note="ORF3"			
/codon_start=1	/codon_start=1			
/transl_table=1	/transl_table=1			
/product="hypothetical protein"	/product="hypothetical protein"			
/protein_id="CAB94378.1"	/protein_id="CAB94378.1"			
/db_xref="GI:8517642"	/db_xref="GI:8517642"			
/db_xref="SPTREMBL:O9LCV1"	/db_xref="SPTREMBL:O9LCV1"			

gene
CDS
/translation="MKPRRLAAACULLLAAAPAAAAEIGDGLHNPMLRPTKDLRED
LAENAAQOKRLLILVEORGCITCAEMHEKVEPDEIDALIRDGIFFVQKMLPGVEVET
DEDFGACAAAVRWGMFTPTMIFLPEEVPARKTAQAQAAVAMMPGAFNGKTTLSALL
TWVRNHHGYENGEHOKFLARELEHKN"
2654, .3127
/gene="soxX"
2654, .3127
/gene="soxX"
/codon_start=1
/transl_table=1
/protein_id="CAB94379.1"
/db_xref="GI:8517643"
/db_xref="SPTREMBL:Q9LCV0"
/translation="MSSLHWASVVAANAIAIPAICETAPKEVYVVEGAVEASLIGAPG
NPBEESVRLMTNALGNVCVACHQIALPDVEFPPTIAPPLDAGDRMTAEOLRGIVANA
KMTFGTGPMPAFYKVDFVVRBGDFSGAAGEPLALPLNMQIIEVVAFLVTLKE"
3168, .3590
/gene="soxY"
3168, .3590
/gene="soxY"
/codon_start=1
/transl_table=1
/protein_id="CAB94380.1"
/db_xref="GI:8517644"
/db_xref="SPTREMBL:Q9LC09"
/translation="MILSRREALMIGEGGLAAALPGKAVASTVDELITAAFTGCAGT
GEGGTLTAPETIAENGTRPIEVKAPGAVAIAMLACGPEPVAFTVFNFGPAADQRAA
TRIRLAOTQDVIYALAKMADGSVVAKQTIVKVTIGGCGS"
3612, .3941
/gene="soxZ"
3612, .3941
/gene="soxZ"
/codon_start=1
/transl_table=1
/protein_id="CAB94381.1"
/db_xref="GI:8517645"
/db_xref="SPTREMBL:Q9LC08"
/translation="MADDAKPRVYVSSAKAGETVYKALISHKMEGORKDADGKLI
PSIINRPTCEINGVNVDAVIDPAVSPTNPFEEFDKADVADGEEFKFTMYDDGGSVED
VKPIIYVA"
3976, .4848
/gene="soxA"
3976, .4848
/gene="soxA"
/codon_start=1
/transl_table=1
/protein_id="CAA55827.2"
/db_xref="GI:8517646"
/db_xref="SPTREMBL:O33434"
/translation="MPRETKTKGTLLAATLGLALAGAAFDAPVEDGHLIETSGSPVEI
VTKRPAFLADTPTITSGWHEFRDSTRDLRDEDPDAVFDVDRGLDKNNAMGAVET
GESCSHOGGSESNAGLRAVAPRDEHTGKLTIMEDVYVNCVTEBMRGLEKGVGSDMM
KDMLSILSOSRGMAVYKIDGPAPAFYERHEKREIYTYTQGLESCANCHDNAGNNTI
RADHLSQOQINGFPYTRLKDSGMVTAQHRFVCGVDTRAETFKAGSDFALELYVAS
RNGNGLSVGVSVRH"
4968, .6662
/gene="soxB"
4968, .6662
/gene="soxB"
/codon_start=1
/transl_table=1
/protein_id="CAA55824.2"
/db_xref="GI:8517647"
/db_xref="SPTREMBL:P72177"
/translation="MTRREFRITQVAMASSALVGSAGFETMSRYVAAQOALITDOLQEDP
DEGNTLLHVDIHIAQLPITYFREPEINIGDAGKGVPHATGAEFQMRGELIAGSDSD
AVALYDPDEFALARGYGRMGMDRAVATVKAIRAPDSILLDGDGTWHSQMSFTLTK
GODANVYVNALGVADMTSHWEMTYGTEVEKEIVETOLKFPGLANIDVEDDEFALFEP
YKIFRGGORIGVIGOAPRYMPIANPKMPEYSFGIREEBHOVDELRAGDYDLVV
VLSHGFPVDRKMGGRVKGIDVILSGTHHDVAVPEPTIIGETIILATSGNCKFEVARDL
DYRDKMMGFRKLIPITSDVITTPDADMAALIDAEAPAFKQLEBRIGTSTSLIXRRG
TEMGSMWIDICALITERDVOIALSPGVKQCTLLPGDALTREDIHNTVSTYGAIVR
TENGSETKTLYLEDAVDNIIFNTDPPYQOGDMVRVGLAIVADYSKPMGNSIAWALAA

```

DREAELPAKSYIVAGMASVNEGTBEGRQIWDVYEAIDIRKIGTVRLQPNISVYTGCV"
6686. .7978
/function="sulfur oxidation"
/note="soxc"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="sulfite dehydrogenase"
/protein_id="CAA55829.1"
/db_xref="GI:1550793"
/db_xref="SPTREMBL:P72178"
/translation="MKDELTKDGDVNGTSRRAFELRCGAVAGALVAGAGIARAGTP
DPIITLPIQPAASEFEAGVAHAPYAGLPPIHESVHKROVEVLTSPSINFTPIHALE
GTTTPGCAFERHHSAGLIELSKODYIRLMINGLPEKPLVTFEDLLRPRRTTAFCC
AANGGEMWGAQLEGCCYITQGMILHNEIYVPLSVLLAEAGVRECKMILAEGGDASS
NGESRFEIETSKYITDLPDGRARKMPWVMDAKSVILSPSPQVIRHGKGLVLTGLAWSON
GRTTRDVSLDGGCKNMVTTARITCOALPKALTRFLDIDDMGSEMLLOSRAVDEGYVO
PTKDALAIRGRNNVYHNNGIOTWVWKADGEVENEIQA"
7962. .9116
/gene="sord"
/feature="sord"
7962. .9116
/gene="sord"
/gene="sord"
/codon_start=1

```

```

Query Match      1.8%; Score 38.6; DB 1; Length 10624;
Best Local Similarity 48.8%; Pled. No. 68;
Matches 104; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

OY 1847 TCGACCTCTTCCGACCAAAACTATGTCGTATGCGACGAGAACGTCGATCCATATCG 1906
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4253 TCGAGGCGCGGATCATCAGAAACACATGCGGGGTTCGCAAAATCGCCGCTCAAGATCC 4194

OY 1907 CTTTCTATGCAACAAACGACATCATCGGGCGGCTAATCGCGGTAGCCCCCTTATTCGATPAAGA 1966
      | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4193 CGCGTGTGTGTGTGCGCAAAAGTGCACGCGCGAATTAATGCTGTGCAACGTATTCGGCCAGA 4134

OY 1967 ACGGCCGCTGTGATCGCTCTTGTTCATGGCAATGGGAAAGCTATGATGGTGCATTCG 2026
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4133 AAGCGCGGCGGTGTGGGCTCTTGATGACAGATCTCACCAGCGGTCCGCTGTGTGTCGATGACC 4074

OY 2027 AGTTTGAACCGCATCTGCAGCGCACATCATCGC 2059
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4073 AGCCCGTCTCGACCGGCGGTCCGGAAGGCCGC 4041

Search completed: October 17, 2002, 21:35:14
Job time : 4282 secs

```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 17, 2002, 19:21:20 ; Search time 249 Seconds
(without alignments)
14748.916 Million cell updates/sec

Title: US-10-008-355-1

Perfect score: 2139
Sequence: 1 atgcaatggaataaataag.....aagagctgaagtgtacttaa 2139

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N.Geneseq_032802:*

1: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1980.DAT:*

2: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1981.DAT:*

3: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1982.DAT:*

4: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1983.DAT:*

5: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1984.DAT:*

6: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1985.DAT:*

7: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1986.DAT:*

8: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1987.DAT:*

9: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1988.DAT:*

10: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1989.DAT:*

11: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1990.DAT:*

12: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1991.DAT:*

13: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1992.DAT:*

14: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1993.DAT:*

15: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT:*

16: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1995.DAT:*

17: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1996.DAT:*

18: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:*

19: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT:*

20: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT:*

21: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:*

22: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:*

23: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:*

24: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	42.2	2.0	1147	23	ABLI4089
2	42.2	2.0	3632	23	ABLI4088
3	40.6	1.9	2868	23	ABLI21554
4	38.8	1.8	4013	23	ABLI18653
5	38.6	1.8	3621	22	AAH48797
6	38.6	1.8	4848	22	AAH48798
7	37.8	1.8	10757	23	ABLI21478
8	37.4	1.7	9845	23	ABLI18652
9	37	1.7	786	22	AAH68978

10	37	1.7	1077	22	AAH68976
11	36.8	1.7	810	22	AAH01671
12	36.8	1.7	349980	22	AAH66431
13	36.4	1.7	1090	21	AAH66137
14	36.4	1.7	1292	22	AAH58976
15	36.4	1.7	1649	21	AAH35391
16	36.4	1.7	1768	22	AAH60762
17	36.2	1.7	345	22	AAH67046
18	36.2	1.7	536	21	AAH43983
19	36.2	1.7	2280	17	AAH36127
20	36.2	1.7	349980	22	AAH68530
21	35.6	1.7	6174	22	AAH81576
22	35.6	1.7	8617	22	AAH81576
23	35.6	1.7	32313	22	AAH81578
24	35.2	1.6	449	20	AAH68139
25	35.2	1.6	4590	22	AAH24065
26	35	1.6	1910	23	AAH85131
27	34.8	1.6	632	21	AAH08951
28	34.8	1.6	5059	20	AAH84332
29	34.6	1.6	639	21	AAH3795
30	34.6	1.6	975	23	ABLI6623
31	34.6	1.6	2975	23	ABLI6622
32	34.6	1.6	5133	23	ABLI09568
33	34.2	1.6	3059	8	AAH70566
34	34	1.6	2759	23	ABLI06895
35	34	1.6	3972	24	AAH97214
36	34	1.6	5837	12	AAH14461
37	34	1.6	5837	14	AAH43966
38	34	1.6	5837	14	AAH43515
39	34	1.6	5837	15	AAH66293
40	34	1.6	5837	17	AAH17546
41	34	1.6	5837	18	AAH70812
42	34	1.6	5837	20	AAH68141
43	34	1.6	5837	20	AAH68141
44	34	1.6	13620	23	ABLI06894
45	34	1.6	236303	22	AAH11614

ALIGNMENTS

RESULT 1	
ID	ABLI4089
ABLI4089	standard; cDNA: 1147 BP.
AC	ABLI4089;
DT	26-MAR-2002 (first entry)
XX	
DE	Drosophila melanogaster expressed polynucleotide SEQ ID NO 36749.
XX	
KW	Drosophila; developmental biology; cell signalling; insecticide;
KW	pharmaceutical; gene; ss.
XX	
OS	Drosophila melanogaster.
XX	
PN	WO200171042-A2.
XX	
PD	27-SEP-2001.
XX	
PF	23-MAR-2001; 2001WO-US09231.
XX	
PR	23-MAR-2000; 2000US-191637P.
XX	
PR	11-JUL-2000; 2000US-0614150.
XX	
PI	(PEKE) PE CORP NY.
XX	
PI	Venter JC, Adams M, Li PWD, Myers EW;
XX	WPI; 2001-656860/75.
DR	P-PSDB; ABB69986.
XX	
PT	New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
PS Claim 1; SEQ ID NO 36749; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB101840-AB116175), expressed DNA
CC sequences (AB101840-AB116175) and the encoded proteins
CC (AB101840-AB116175).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1147 BP; 257 A; 335 C; 303 G; 252 T; 0 other;
Query Match 2.0%; Score 42.2; DB 23; Length 1147;
Best Local Similarity 49.3%; Pred. No. 0.085;
Matches 110; Conservative 0; Mismatches 113; Indels 0; Gaps 0;
QY 1470 GGTTCCTTATAGCGACAGTTCATGCCATGCTCAAGTCCATGACAGAAAGTTTGC 1529
DB 711 GGCCGCTATACGCCGCTGCTGCTGCCCTCCAGAAATCCAGCAGCTGCTTGT 770
QY 1530 CAAGCCTATCGAAGAAAGATCCGCGAGTAGAGCTTCCAAAGAGCTAATAGCTGCTCG 1589
DB 771 CCAACACACGAGAAAGAGCGTGGCAGATCGCGTCCAGAGGTATACCATTTGCCCTTC 830
QY 1590 CGCTATTACGCCGATCGATGCCCAATGCCATGCTATGCCATTGAGAAGGCAAGCCTTTT 1649
DB 831 CACCAACCAATCCCTACCAACCTGCGCGTCTACTAGGCGACGACATCCCGGATCTTCAA 890
QY 1650 CTTTGGCGGTTTGGCTGAGATGTACCCCGAGCGTCTGCGC 1692
DB 891 CATCATCCTGTGTTCATGTCGTCTTCGACGTCTCTGCTG 933
RESULT 2
AB114088
ID ABL14088 standard; cDNA; 3632 BP.
XX
AC ABL14088;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 36746.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI: 2001-656860/75.
XX
PT P-PSDB; ABB69985.
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions -
XX
PS Claim 1; SEQ ID NO 36746; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB101840-AB116175), expressed DNA
CC sequences (AB101840-AB116175) and the encoded proteins
CC (AB101840-AB116175).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 3632 BP; 1062 A; 799 C; 773 G; 998 T; 0 other;
Query Match 2.0%; Score 42.2; DB 23; Length 3632;
Best Local Similarity 49.3%; Pred. No. 0.16;
Matches 110; Conservative 0; Mismatches 113; Indels 0; Gaps 0;
QY 1470 GGTTCCTTATAGCGACAGTTCATGCCATGCTCAAGTCCATGACAGAAAGTTTGC 1529
DB 2196 GGCCGCTATACGCCGCTGCTGCTGCCCTCCAGAAATCCAGCAGCTGCTTGT 2255
QY 1530 CAAGCCTATCGAAGAAAGATCCGCGAGTAGAGCTTCCAAAGGCTAATAGCTGCTCG 1589
DB 2256 CCAACACACGAGAAAGAGCGTGGCAGATCGCGTCCAGAGGTATACCATTTGCCCTTC 2315
QY 1590 CGCTATTACGCCGATCGATGCCCAATGCCATGCTATGCCATTGAGAAGGCAAGCCTTTT 1649
DB 2316 CACCAACCAATCCCTACCAACCTGCGCGTCTACTAGGCGACGACATCCCGGATCTTCAA 2375
QY 1650 CTTTGGCGGTTTGGCTGAGATGTACCCCGAGCGTCTGCGC 1692
DB 2376 CATCATCCTGTGTTCATGTCGTCTTCGACGTCTCTGCTG 2418
RESULT 3
AB121554
ID ABL21554 standard; DNA; 2868 BP.
XX
AC ABL21554;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 16135.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI: 2001-656860/75.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX

PS	Claim 1, SEQ ID NO 16135; 21np + Sequence Listing; English.
XX	
CC	The invention relates to an isolated nucleic acid detection reagent
CC	capable of detecting 1000 or more genes from <i>Drosophila</i> . The invention is
CC	useful in developmental biology and in elucidating cell signalling and
CC	cell-cell interactions in higher eukaryotes for the development of
CC	insecticides, therapeutics and pharmaceutical drugs. The invention
CC	discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC	sequences (AB101840-AB16175) and the encoded proteins
CC	(AAB57737-AB17022).
CC	The sequence data for this patent did not form part of the printed
CC	specification, but was obtained in electronic format directly from WIPO
CC	at ftp.wipo.int/pub/published_npt_sequences.
XX	
SO	Sequence 2868 BP; 737 A; 657 C; 624 G; 850 T; 0 other;
	Query Match 1.9%; Score 40.6; DB 23; Length 2868;
	Best Local Similarity 50.3%; Pred. No. 0.41;
	Matches 100; Conservative 0; Mismatches 99; Indels 0; Gaps 0;
OY	1494 TCCATGCTCAAGTCCATGACAGAGAAAAGTTGCCAAGGCTATTCAGAGAAATCCGGC 1553
DB	10 TCCCTCCCAAGAGTCCACGACGACTCGTTCTGTTGTCACAGACACGAGAGAGAGCTGAC 69
OY	1554 AGTAAGCTTTCCAGACGCGTAAATAGCTGCTCTGCGCTATTCAGAGCCGATCCATGGC 1613
DB	70 AGCATTCGGCGTCCAGCGTGATGCCCTTCACACCAATCCCTACAACTGGC 129
OY	1614 CAATGCTATGCGCATTTAGAGAAAGGCGAGCGCTTTTCTTGCCGGTTGCGTAGATGTA 1673
DB	130 CGTCTACTACGCGACGACGACTACCGGATGATCTTCAACATCATCCTGTGTTATGTTGCT 189
OY	1674 CCCCAGCGTGGCTTGCCG 1692
DB	190 CTTCGAGCTGCTCTGCTG 208
	RESULT 4
AB18653	
ID	AB18653 standard; DNA; 4013 BP.
XX	
AC	AB18653;
XX	
DT	26-MAR-2002 (first entry)
XX	
DE	<i>Drosophila melanogaster</i> genomic polynucleotide SEQ ID NO 7432.
XX	
KW	<i>Drosophila</i> ; developmental biology; cell signalling; insecticide;
XX	Pharmaceutical; gene; ds.
OS	<i>Drosophila melanogaster</i> .
XX	
PN	WO2001/1042-A2.
XX	
PD	27-SEP-2001.
XX	
PE	23-MAR-2001; 2001WO-US09231.
XX	
PR	23-MAR-2000; 2000US-191637P.
XX	
PR	11-JUL-2000; 2000US-0614150.
XX	
PA	(PEKE) PE CORP NY.
XX	
PI	Venter JC, Adams M, Li PWD, Myers EW;
XX	
DR	WPI; 2001-656860/75.
XX	
PT	New isolated nucleic acid detection reagent for detecting 1000 or more
XX	genes from <i>Drosophila</i> and for elucidating cell signalling and cell-cell
PS	interactions -
XX	
PS	Claim 1; SEQ ID NO 7432; 21np + Sequence Listing; English.
XX	

[illegible]

```
FT      /product= "soxa"
XX
XX      DE1006201-A1.
XX
XX      16-AUG-2001.
XX
XX      11-FEB-2000; 2000DE-1006201.
XX
XX      11-FEB-2000; 2000DE-1006201.
XX
XX      (FRIE/) FRIEDRICH C.
XX
XX      Friedrich C;
XX
XX      WPI: 2001-497622/55.
XX
XX      P-PSDB: AAB86529, AAB86530, AAB86531, AAB86532, AAB86533, AAB86534.
XX
XX      New genes involved in oxidation of reduced sulphur compounds,
XX      particularly conversion of thiosulphate to sulphate, and related
XX      proteins, from Paracoccus pantotrophus -
XX
XX      Claim 1; Page 10-14; 22pp: German.
XX
XX      This invention describes novel genes (I) of the sox (sulphur oxidation)
XX      region designated ORF1, ORF2, ORF3 (Open Reading Frame), and soxXYZA
XX      from Paracoccus pantotrophus GB17. The proteins (II) encoded by (I) are
XX      used for oxidation of reduced sulphur compounds in cellular or cell-free
XX      systems, particularly of thiosulphate to sulphate. (I) is also used for
XX      expression of the proteins and to prepare transgenic plants or animals
XX      or transformed microorganisms. This sequence encodes the ORF2, ORF3 and
XX      the soxXYZA proteins from the P. pantotrophus sox region.
XX
XX      Sequence 3621 BP; 650 A; 1153 C; 1186 G; 632 T; 0 other;
SQ
Query Match      1.8%; Score 38.6; DB 22; Length 3621;
Best Local Similarity 48.8%; Pred. No.1.8;
Matches 104; Conservative 0; Mismatches 109; Indels 0; Gaps 0;
OY      1847 TCGACCTCTTCGCGACCAAAACATGTCGCTATGCCGAGACGGTGCCTCATATCG 1906
        |||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      3026 TCGAGCGCGCGATCCACGAAACATGCGGGGCTTGCGAATGCTCCCGCTCAAGATCC 2967
OY      1907 CTTTCCTATCGAACACGACATCAGCGGGGTAACCTCCGGTAGCCCCGTATTGATAAGA 1966
        |||| | | | | | | | | | | | | | | | | | | | | | |
DB      2966 CGCGTGTCTGTCGCGGAAAGTCCAGCCGGAATGATGTGCAACGTATCGGCCAGA 2907
OY      1967 ACGGCGCTGTAGTGGTCTGCTTTCGATGGCACTGGGAAGCTATGAGTGTGATCAG 2026
        |||| | | | | | | | | | | | | | | | | | | | | | |
DB      2906 AAGCGGGGCGGTGCGCTTGCTCAGATCTTCGACCGGTCCGCTGCTGATGAC 2847
OY      2027 AGTTGAACCCGATCTGCAGCGACACATCAGCG 2059
        || | | | | | | | | | | | | | | | | | | | |
DB      2846 AGCCGCTCTGACCGGCTCCGCGAAGGCCGCG 2814

RESULT 6
AAH48798/C
ID      AAH48798 standard; DNA; 4848 BP.
XX
XX      AAH48798;
XX
XX      05-NOV-2001 (first entry)
XX
XX      P. pantotrophus GB17 DNA encoding sox-associated proteins.
XX
XX      Sulphur oxidation; sox; ORF1; ORF2; ORF3; soxX; soxY; soxZ; soxA;
XX      reduced sulphur; thiosulphate; transgenic plant; transgenic animal; ds.
XX
XX      Paracoccus pantotrophus.
XX
XX      DE1006201-A1.
XX
XX      16-AUG-2001.
XX
```

```
XX
XX      11-FEB-2000; 2000DE-1006201.
XX
XX      11-FEB-2000; 2000DE-1006201.
XX
XX      (FRIE/) FRIEDRICH C.
XX
XX      Friedrich C;
XX
XX      WPI: 2001-497622/55.
XX
XX      New genes involved in oxidation of reduced sulphur compounds,
XX      particularly conversion of thiosulphate to sulphate, and related
XX      proteins, from Paracoccus pantotrophus -
XX
XX      Claim 1; Fig 1; 22pp: German.
XX
XX      This invention describes novel genes (I) of the sox (sulphur oxidation)
XX      region designated ORF1, ORF2, ORF3 (Open Reading Frame), and soxXYZA
XX      from Paracoccus pantotrophus GB17. The proteins (II) encoded by (I) are
XX      used for oxidation of reduced sulphur compounds in cellular or cell-free
XX      systems, particularly of thiosulphate to sulphate. (I) is also used for
XX      expression of the proteins and to prepare transgenic plants or animals
XX      or transformed microorganisms. This sequence encodes the P. pantotrophus
XX      sox region associated proteins described in the invention.
XX
XX      Sequence 4848 BP; 885 A; 1584 C; 1552 G; 827 T; 0 other;
SQ
Query Match      1.8%; Score 38.6; DB 22; Length 4848;
Best Local Similarity 48.8%; Pred. No.2.1;
Matches 104; Conservative 0; Mismatches 109; Indels 0; Gaps 0;
OY      1847 TCGACCTCTTCGCGACCAAAACATGTCGCTATGCCGAGACGGTGCCTCATATCG 1906
        |||| | | | | | | | | | | | | | | | | | | | | | |
DB      4253 TCGAGCGCGCGATCCACGAAACATGCGGGGCTTGCGAATGCTCCCGCTCAAGATCC 4194
OY      1907 CTTTCCTATCGAACACGACATCAGCGGGGTAACCTCCGGTAGCCCCGTATTGATAAGA 1966
        |||| | | | | | | | | | | | | | | | | | | | | | |
DB      4193 CGCGTGTCTGTCGCGGAAAGTCCAGCCGGAATGATGTGCAACGTATCGGCCAGA 4134
OY      1967 ACGGCGCTGTAGTGGTCTGCTTTCGATGGCACTGGGAAGCTATGAGTGTGATCAG 2026
        |||| | | | | | | | | | | | | | | | | | | | | | |
DB      4133 AAGCGGGGCGGTGCGCTTGCTCAGATCTTCGACCGGTCCGCTGCTGATGAC 4074
OY      2027 AGTTGAACCCGATCTGCAGCGACACATCAGCG 2059
        || | | | | | | | | | | | | | | | | | | | |
DB      4073 AGCCGCTCTGACCGGCTCCGCGAAGGCCGCG 4041

RESULT 7
ABL21478
ID      ABL21478 standard; DNA; 10757 BP.
XX
XX      ABL21478;
XX
XX      26-MAR-2002 (first entry)
XX
XX      Drosophila melanogaster genomic polynucleotide SEQ ID NO 15907.
XX
XX      Drosophila; developmental biology; cell signalling; insecticide;
XX      pharmaceutical; gene; ds.
XX
XX      Drosophila melanogaster.
XX
XX      W0200171042-A2.
XX
XX      23-MAR-2001; 2001WO-US09231.
XX
XX      23-MAR-2000; 2000US-191637P.
XX
XX      11-JUL-2000; 2000US-0614150.
XX
```

PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI: 2001-656860/75.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 15907; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 10757 BP; 3125 A; 2268 C; 2226 G; 3138 T; 0 other;
XX
Query Match 1.8%; Score 37.6; DB 23; Length 10757;
Best Local Similarity 54.1%; Pred. No. 5.4;
Matches 100; Conservative 0; Mismatches 82; Indels 3; Gaps 1;
XX
OY 1283 AATGCGCTGACGACAGTACAAAGACTACCCCTCGCTGACGCTAAGGTCGCGCG 1342
DB 6709 ATTGTTTTCGACATGACCAAGTCTCGATGCCATTTGAGGAAAGTTCAAGCCGG 6768
OY 1343 CCATGCTGATATTTGACCGCGGTATCCCTGCCAGACCTCCCGATATATTCAGA 1402
DB 6769 GT---GACTTATATCATTCGCGGATGACGAGGATTAATTTATTTGAATGT 6825
OY 1403 ATGTAATGACAGAAATTCAAAGCGCACGACGAGATATGACACTTCGATTCGACA 1462
DB 6826 AAGTAAACAAACAAATTCATCAATCAATCAATTAATGACACCTTCTGTTTGCA 6885
OY 1463 AGAGT 1467
DB 6886 AAAAT 6890
XX
RESULT 8
AB18652/c
ID ABL18652 standard; DNA; 9845 BP.
XX
AC ABL18652;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 7429.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX
PR 11-JUL-2000; 2000US-061415O.
XX
PA (PEKE) PE CORP NY.
XX

PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI: 2001-656860/75.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 7429; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 9845 BP; 2751 A; 2130 C; 2204 G; 2760 T; 0 other;
XX
Query Match 1.7%; Score 37.4; DB 23; Length 9845;
Best Local Similarity 49.7%; Pred. No. 6.8;
Matches 95; Conservative 0; Mismatches 96; Indels 0; Gaps 0;
XX
OY 1294 GACAAATCAAAAGACTACCTCCCTCGCTGACGCTAGTGTGCCCATGCTCGAT 1353
DB 8624 GACATATATAGGAGTGTGCTCATTTGATGTGCGTATATTCGCGGTGCGATGCC 8565
OY 1354 AFTGTAGCCGCGGTATCCCTGCCAGCAAGCTCCCGATATATTCAGATGTAATGAC 1413
DB 8564 AFTGGTCAAGACCTTGTGCTGACCAACCGCGGACCAACCGCATACATCAT 8505
OY 1414 AAGAATTCAAAGGAGACAGCAAGATGACAGCTCGATTCGACAGAGTGTGT 1473
DB 8504 CGAATGCTTTCCTGCGCTTCAAGCATGATCTGAGCTGATTCGACGACACTTCTGT 8445
OY 1474 CCTATACGA 1484
DB 8444 CCCGAGGCGCA 8434
XX
RESULT 9
AAC68978
ID AAC68978 standard; DNA; 786 BP.
XX
AC AAC68978;
XX
DT 27-FEB-2001 (first entry)
XX
DE Merozoite surface protein-133 coding sequence.
XX
KW Merozoite surface protein; protozoacide; vaccine; malaria; ss.
XX
OS Plasmodium falciparum.
XX
PN WO200063245-A2.
XX
PD 26-OCT-2000.
XX
PF 20-APR-2000; 2000WO-GB01558.
XX
PF 20-APR-1999; 99GB-0009072.
XX
PR 13-MAY-1999; 99US-0311817.
XX
PR 25-MAY-1999; 99CA-2271451.
XX
PA (MEDT-) MEDICAL RES COUNCIL.
XX
PI Holder A, Birdsell B, Feeney J, Morgan W, Syed S, Uthairibull C;
XX

DR WPI; 2001-015762/02.
 DR P-PSDB; AAB37610.
 XX Novel variants of the C-terminal fragment of Plasmodium merozoite
 PT surface protein-1, useful as vaccines for treating or preventing
 PT malaria -
 XX
 PS Example 5; Fig 15; 126pp; English.
 XX
 CC The present invention relates to non-natural variants of a C-terminal
 CC fragment of a Plasmodium merozoite surface protein-1 (MSP-1). The
 CC non-natural variants have reduced affinity for at least 1 antibody
 CC capable of blocking a second antibody that inhibits the proteolytic
 CC cleavage of Plasmodium MSP-1.4.2, and has the same affinity for at least
 CC one third antibody that inhibits the proteolytic cleavage of Plasmodium
 CC MSP-1.4.2, compared to natural MSP-1.1.9. The non-natural variants of the
 CC present invention are useful for immunising a mammal against malaria, and
 CC can be used to treat malaria. The present sequence is the MSP-133
 CC coding sequence.
 XX
 SQ Sequence 786 BP; 245 A; 164 C; 159 G; 218 T; 0 other;
 Query Match 1.7%; Score 37; DB 22; Length 786;
 Best Local Similarity 52.2%; Pred. No. 2.4; Mismatches 0; Gaps 0;
 Matches 82; Conservative 0; Indels 75; Indels 0;
 QY 1377 CGACACGCTCCCGATATATTCAGAAATGATCGACAGAAATTCAGAGCGACAGAA 1436
 DB 135 CGTTACGTCAGAGACATTTTGAACCTCCAGATTCACAGAGAGAGAAACTTCAGAAACGT 194
 QY 1437 GAAGTATGACAGCTTCGTATTCGACAAAGAGTGTGCTTATAGCGACAGATTCATGC 1496
 DB 195 TCTGAGCTGTGACTTGATTCATCAAGAGATTTGACTTCTTAACGTTGTTAAGA 254
 QY 1497 CATGCTCAGTCATGACAGAAAGATTTGCCAAG 1533
 DB 255 CCATACAGATTCCTGAAACAGAGAGAGACAG 291
 RESULT 10
 AAC68976
 ID AAC68976 standard; DNA: 1077 BP.
 XX
 AC AAC68976;
 XX
 DT 27-FEB-2001 (first entry)
 XX
 DE Merozoite surface protein-142 coding sequence.
 XX
 KM Merozoite surface protein; protazoicide; vaccine; malaria; ss.
 XX
 OS Plasmodium falciparum.
 XX
 PN WO200063245-A2.
 XX
 PD 26-OCT-2000.
 XX
 PF 20-APR-2000; 2000WO-GB01558.
 XX
 PR 20-APR-1999; 99GB-0009072.
 PR 13-MAY-1999; 99US-0311817.
 PR 25-MAY-1999; 99CA-2271451.
 XX
 PA (MEDIC-) MEDICAL RES COUNCL.
 XX
 PI Holder A, Birdsall B, Feeney J, Morgan W, Syed S, Uthapibull C;
 DR WPI; 2001-015762/02.
 XX
 PT Novel variants of the C-terminal fragment of Plasmodium merozoite
 PT surface protein-1, useful as vaccines for treating or preventing
 PT malaria -
 XX

PS Example 5; Fig 15; 126pp; English.
 XX
 CC The present invention relates to non-natural variants of a C-terminal
 CC fragment of a Plasmodium merozoite surface protein-1 (MSP-1). The
 CC non-natural variants have reduced affinity for at least 1 antibody
 CC capable of blocking a second antibody that inhibits the proteolytic
 CC cleavage of Plasmodium MSP-1.4.2, and has the same affinity for at least
 CC one third antibody that inhibits the proteolytic cleavage of Plasmodium
 CC MSP-1.4.2, compared to natural MSP-1.1.9. The non-natural variants of the
 CC present invention are useful for immunising a mammal against malaria, and
 CC can be used to treat malaria. The present sequence is the MSP-142
 CC coding sequence.
 XX
 SQ Sequence 1077 BP; 338 A; 228 C; 225 G; 286 T; 0 other;
 Query Match 1.7%; Score 37; DB 22; Length 1077;
 Best Local Similarity 52.2%; Pred. No. 2.8; Mismatches 0; Gaps 0;
 Matches 82; Conservative 0; Indels 75; Indels 0;
 QY 1377 CGACACGCTCCCGATATATTCAGAAATGATCGACAGAAATTCAGAGCGACAGAA 1436
 DB 135 CGTTACGTCAGAGACATTTTGAACCTCCAGATTCACAGAGAGAGAAACTTCAGAAACGT 194
 QY 1437 GAAGTATGACAGCTTCGTATTCGACAAAGAGTGTGCTTATAGCGACAGATTCATGC 1496
 DB 195 TCTGAGCTGTGACTTGATTCATCAAGAGATTTGACTTCTTAACGTTGTTAAGA 254
 QY 1497 CATGCTCAGTCATGACAGAAAGATTTGCCAAG 1533
 DB 255 CCATACAGATTCCTGAAACAGAGAGAGACAG 291
 RESULT 11
 AAH01671/C
 ID AAH01671 standard; DNA: 810 BP.
 XX
 AC AAH01671;
 XX
 DT 24-JUL-2001 (first entry)
 XX
 DE Fonsecaea pedrosol nucleotide sequence SEQ ID NO:1664.
 XX
 KM Species specific; genus specific; family specific; probe; detection;
 KM identification; algal; archaeal; bacterial; fungal; parasitical;
 KM microorganism; diagnosis; translation elongation factor Tu; toxin;
 KM translation elongation factor G; RecA recombinase; resistance;
 KM catalytic subunit of proton-translocating ATPase; antimicrobial;
 KM vaccine; primer; ds.
 XX
 OS Fonsecaea pedrosol.
 XX
 PN WO200123604-A2.
 XX
 PD 05-APR-2001.
 XX
 PF 28-SEP-2000; 2000WO-CA01150.
 XX
 PR 28-SEP-1999; 99CA-2283458.
 PR 19-MAY-2000; 2000CA-2307010.
 XX
 PA (INFE-) INFECTIO DIAGNOSTIC (IDI) INC.
 XX
 PI Bergeron MG, Boissinot M, Huletsky A, Menard C, Ouellette M;
 PI Picard FU, Roy PH.
 DR WPI; 2001-245006/25.
 XX
 PT Nucleic acid sequences are used to generate universal probes and
 PT primers which can be used to identify and detect the presence of algal,
 PT archaeal, bacterial, fungal and parasitical species in a test sample -
 XX
 XX Claim 24; Page 1264-1265; 1580pp; English.

Query Match	1.7%	Score 36.8	DB 22	Length 810
Best Local Similarity	46.2%	Pred. No. 2.8		
Matches 122; Conservative	0	Mismatches 142	Indels 0	Gaps 0

RESULT 12	
AAF86431/C	
ID AAF86431	standard; DNA; 349980 BP.

FH	Key	Location/Qualifiers
FT	misc_feature	300001..349980

FT /note="This sequence overlaps with the 5' end of
FT AAH41223"

PS Claim 1; Page 183-279; 1657pp; French.

CC AAH75903-AAH75920 and AAG66436.

Matches 95; Conservative 0; Mismatches

QY 1557 AGAGCTTTCCAA 1568
||| |
Db 167984 CCAGCTCGCCGA 167973

DT 13-FEB-2001 (first entry)

DE Human kidney disease associated gene SEQ ID 7.

KW Nephrotropic: cytostatic; human; kidney disease associated; hypertension
KW Batten's syndrome; Gitelman syndrome; nephrolithiasis; renal failure
KW renal amyloidosis: primary aldosteronism; Addison's disease; cancer;
KW glomerulonephritis; dysplastic malformation; medullary cystic disease;
KW medullary sponge kidney; tubular dysplasia; Alport's syndrome; ds.
XX Homo sapiens.
XS

OS Homo sapiens.

XX WO20061622-A2.
XX 19-OCT-2000.
XX PD
XX PE 28-MAR-2000; 2000WO-US08260.
XX PR 09-APR-1999; 99US-0289349.
XX PA (INCYTE) INCYTE PHARM INC.
XX PI Walker MG, Volkmuth W, Klingler TM, Azimzal Y, Yue H;
XX DR WPI: 2000-665116/64.
XX DR P-PSDB: AAY85680.
XX PT Novel kidney disease associated gene useful for diagnosing treating and
XX PT preventing renal disorders, is utromodulin, NKCC2, NCC2, aldolase B,
XX PT ROMK1, ATP1G1, PDZK1, NPT1, calbindin, kininogen or CIC-Kb
XX PS Claim 2; Page 33-34; 36pp; English.
XX CC Polynucleotides AAC66131-C66139 represent human kidney disease
XX CC associated genes. Proteins AA185678-Y85680 represent human kidney disease
XX CC associated proteins. The polynucleotides have nephrotropic and cytoskeletal
XX CC activity. The polynucleotides, encoded proteins and pharmaceutical
XX CC compositions containing them are useful for diagnosing, prognosing,
XX CC treating and preventing renal disorders such as Bartter's syndrome,
XX CC Gitelman syndrome, autosomal dominant polycystic kidney disease and
XX CC nephrolithiasis. The genes and proteins are also useful for identifying
XX CC biomolecules that are associated with a specific disease, regulatory
XX CC pathway, subcellular compartment, cell type, tissue type or species,
XX CC which is useful in diagnosis, prognosis, treatment and evaluation of
XX CC therapies for renal diseases for e.g. renal amyloidosis, hypertension,
XX CC primary aldosteronism, Addison's disease, renal failure,
XX CC glomerulonephritis, chronic glomerulonephritis, tubulointerstitial
XX CC nephritis, cystic disorders and dysplastic malformations, inherited
XX CC polycystic renal disease (PRD), medullary cystic disease, medullary
XX CC sponge kidney and tubular dysplasia, Alport's syndrome, non-renal
XX CC cancers, kidney adenocarcinoma, metastatic renal carcinoma, multiple
XX CC myeloma and nephrotoxic disorders.
XX SQ Sequence 1090 BP; 215 A; 364 C; 303 G; 208 T; 0 other;
Query Match 1.7%; Score 36.4; DB 21; Length 1090;
Best Local Similarity 49.5%; Pred. No. 4.2;
Matches 94; Conservative 0; Mismatches 96; Indels 0; Gaps 0;
OY 1674 CCCCAGACGTCCTCTGCGAGCGATGCCAATTCACATGCGTATGAGCTACGGCTCCAT 1733
DB 136 CCTGACCGCTGTCTTTCACACCTACCAAGCTCATGCACACGACAGCTGACTTCGT 195
OY 1734 CAAGGATATGAACCGCAGACGCGCTGTACACATATATACAGCAGAGGGCGT 1793
DB 196 CAGGAGCAAGCATGCCAGTTTGGGGGCTTCTCTACAAAGAAATGACAGTCAGGAGC 255
OY 1794 ATTGGAGAGCAGATCTTAAGACGATGTTGCCGTACAGAGATATCTCGACCT 1853
DB 256 CGTGACCTGCTGGATGGCGTGGTGATGATCGAGCCGACCGATATTTCCCAATC 315
OY 1854 CTTCCGACAC 1863
DB 316 CTTCCATGCC 325
RESULT 14
AA158976
ID AA158976 standard; cDNA: 1292 BP.
AC AA158976;
XX 22-OCT-2001 (first entry)
XX

DE Human polynucleotide SEQ ID NO 1179.
XX Human; noctropic; immunosuppressant; cytoskeletal; gene therapy; cancer;
XX peripheral nervous system; neuropathy; central nervous system; CNS;
XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
XX chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
XX leukemia; ss.
XX OS Homo sapiens.
XX PN WO200153312-A1.
XX PD 26-JUL-2001.
XX PE 26-DEC-2000; 2000WO-US34263.
XX PR 21-JAN-2000; 2000US-0488725.
XX PR 25-APR-2000; 2000US-0552317.
XX PR 09-JUL-2000; 2000US-0598042.
XX PR 19-JUL-2000; 2000US-0620312.
XX PR 03-AUG-2000; 2000US-0653450.
XX PR 14-SEP-2000; 2000US-0662191.
XX PR 19-OCT-2000; 2000US-0693036.
XX PR 29-NOV-2000; 2000US-0727344.
XX PA (HYSEQ) HYSEQ INC.
XX PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX PI Wang Z, Wang Z, Weinman T, Xu C, Xue AJ, Yang Y, Zhang J;
XX PI Zhao QA, Zhou P, Goodrich R, Drmanac R;
XX DR WPI: 2001-442253/47.
XX DR P-PSDB: AAM39820.
XX PT Novel nucleic acids and polypeptides, useful for treating disorders
XX PT such as central nervous system injuries -
XX PS Claim 1; SEQ ID NO 1179; 10078pp; English.
XX CC The invention relates to human nucleic acids (AA157798-AA161369) and
XX CC the encoded polypeptides (AA158642-AA162213) with noctropic,
XX CC immunosuppressant and cytoskeletal activity. The polynucleotides are useful
XX CC in gene therapy. A composition containing a polypeptide or polynucleotide
XX CC of the invention may be used to treat diseases of the peripheral nervous
XX CC system, such as peripheral nervous injuries, peripheral neuropathy and
XX CC localized neuropathies and central nervous system diseases, such as
XX CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX CC utilisation of the activities such as: Immune system suppression,
XX CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
XX CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
XX CC assays for receptor activity, arthritis and inflammation, leukaemia and
XX CC C.N.S disorders.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification.
XX SQ Sequence 1292 BP; 254 A; 422 C; 368 G; 248 T; 0 other;
Query Match 1.7%; Score 36.4; DB 22; Length 1292;
Best Local Similarity 49.5%; Pred. No. 4.6;
Matches 94; Conservative 0; Mismatches 96; Indels 0; Gaps 0;
OY 1674 CCCCAGACGTCCTCTGCGAGCGATGCCAATTCACATGCGTATGAGCTACGGCTCCAT 1733
DB 339 CCTGACCGCTGTCTTTCACACCTACCAAGCTCATGCACACGACGACGACTTCGT 398
OY 1734 CAAGGATATGAACCGCAGACGCGTGTGTACACATATATACAGCAGAGGGCGT 1793
DB 399 CAGGAGCAAGCATGCCAGTTTGGGGGCTTCTCTACAAAGAAATGACAGTCAGGAGC 458
OY 1794 ATTGGAGAGCAGATCTTAAGACGCGATGATGTTGCCGTACAGAGATATCTCGACCT 1853
DB 458 CTTCCATGCC 463

Dd	459	CGTGGACCTGCTGATGAGGCTGGTGATGATGATGCAGACCAGCAGTAGATTTCGCCACTC	518
Oy	1854	CTTCCGCAC	1863
Db	519	CTTCCATGCC	528
RESULT 15			
ID	AAZ35391	standard; cDNA; 1649 BP.	
XX	AAZ35391;		
AC			
XX			
Dt	11-APR-2000	(first entry)	
XX			
DE	Human kidney-specific protein KR86 cDNA.		
KM	KR86; human; renal failure; kidney disease; glomerulonephritis;		
KW	hypertension; polycystic renal disease; antidiabetic; nephrotropic;		
RN	diagnosis; gene therapy; vaccine; ss.		
XX			
OS	Homo sapiens.		
XX			
FH	Key	Location/Qualifiers	
FT	CDS	578..1435	
FT		/tag= a	
XX			
PN	WO964435-A1.		
PD	16-DEC-1999.		
XX			
PF	27-MAY-1999;	99WO-US11757.	
PR	10-JUN-1998;	98US-0095627.	
XX			
PA	(SMK) SMITHKLINE BEECHAM CORP.		
XX			
Pt	Chen Z, Fredrickson T, Hu E;		
DR	WPt: 2000-116519/10.		
DR	P-PSDB; AAT58646.		
XX			
Pt	New human kidney specific gene KR86 polypeptides, used to identify		
Pt	agonists, antagonists and inhibitors for use in therapy -		
XX			
PS	Claim 2; Page 27; 33pp; English.		
XX			
CC	The present sequence represents cDNA coding for human KR86, a novel		
CC	protein whose expression is highly restricted to the kidney, and		
CC	which may play a role in kidney-related diseases. The invention		
CC	provides polynucleotides having at least 70% identity with the		
CC	present sequence, as well as KR86 polypeptides, recombinant		
CC	materials and methods for their production. KR86 polypeptides may		
CC	be used for identifying agonists and antagonists/inhibitors, and		
CC	for detecting diseases associated with inappropriate KR86 activity		
CC	or levels. KR86 polypeptides and polynucleotides, agonists,		
CC	antagonists and antibodies are used to treat kidney-related		
CC	diseases, such as acute or chronic renal failure, diabetes, kidney		
CC	disease (including glomerulonephritis), hypertension, and polycystic		
CC	renal disease. The polynucleotides are also useful as sources of		
CC	primers and probes which can be used to identify mutations in the		
CC	KR6 gene, and also for detecting the above diseases.		
XQ	Sequence 1649 BP; 376 A; 474 C; 436 G; 363 T; 0 other;		
XX			

Query Match	1.7%;	Score 36.4;	DB 21;	Length 1649;
Best Local Similarity	49.5%;	Pred. No. 5.2;		
Matches 94;	Conservative 0;	Mismatches 96;	Indels 0;	Gaps 0;

QY 1674 CCCCCGAGCGTCTTGCCGAGGATGCCACTTCACCATGCGTATGAGCTACGGCTCCAT 17333
 ||| ||||| - - - - - ||||| - - - - -
 Db 685 CCGGAGCCGTCTTCACCACTCAAGCTCATGACACGACACGAGCAGTGGACTCGT 744

QY 1734 CAAGGATATGAAACCGACGAGGAGGTGCTTGTAACAATCTATCATACGACGAGCAAGGGGCT 1793

Db 745 CAGGAGCAAGCATGCCCAAGTTTGGGGGCTTCTCTCAAGAAATATGACAGTCATGGAGGC 804

QY 1794 ATTGGACAGACGAGTCTCTAAGACCATGATGTTGCCGTACAGGAAATATCTCTGACCT 1853

Db 805 CGTGGACCTGCTGGATGGGCGCTGTGGATGATGCGACCGGAGTGAATTTCCCAACTC 864

QY 1854 CTTCGCGACC 1863

Db 865 CTTCGATGCC 874

Search completed: October 17, 2002, 20:33:09
Job time : 828 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 17, 2002, 20:13:20 ; Search time 60 seconds

(without alignments) updates/sec
8756.834 Million cell

Title: US-10-008-355-1

Perfect score: 2139

Sequence: 1 atgcaatgaatcaaaaag.....aagagctgaagttgatctaa 2139

Scoring table:

IDENTITY_MUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_NA: *
1: /cgn2_6/ptodata/2/lna/5A.COMB.seq: *
2: /cgn2_6/ptodata/2/lna/5B.COMB.seq: *
3: /cgn2_6/ptodata/2/lna/6A.COMB.seq: *
4: /cgn2_6/ptodata/2/lna/6B.COMB.seq: *
5: /cgn2_6/ptodata/2/lna/PCTUS.COMB.seq: *
6: /cgn2_6/ptodata/2/lna/Backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	36.4	1.7	1090	US-09-289-349-7	Sequence 7, Appl
2	36.2	1.7	2280	US-08-415-818-5	Sequence 5, Appl
3	36.2	1.7	2280	US-08-894-236-5	Sequence 5, Appl
4	36.2	1.7	2280	PCT-US96-01444-5	Sequence 5, Appl
5	36	1.7	7218	US-08-232-463-14	Sequence 14, Appl
6	35.2	1.6	449	US-08-825-782-2	Sequence 2, Appl
7	34	1.6	5837	US-07-686-340-1	Sequence 1, Appl
8	34	1.6	5837	US-08-004-139B-1	Sequence 1, Appl
9	34	1.6	5837	US-08-117-491-1	Sequence 1, Appl
10	34	1.6	5837	US-08-271-364A-1	Sequence 1, Appl
11	34	1.6	5837	US-08-811-492-1	Sequence 1, Appl
12	34	1.6	5837	US-08-222-715B-1	Sequence 1, Appl
13	34	1.6	5837	PCT-US96-10545A-1	Sequence 1, Appl
14	33.6	1.6	50937	US-09-773-816-1	Sequence 1, Appl
15	33.6	1.6	50937	US-09-428-517-1	Sequence 1, Appl
16	33.2	1.6	939	US-09-105-390-45	Sequence 45, Appl
17	33.2	1.6	1020	US-09-105-390-61	Sequence 61, Appl
18	33.2	1.6	2612	US-09-105-390-7	Sequence 7, Appl
19	32.4	1.5	710	US-08-998-416-603	Sequence 603, App
20	32.2	1.5	3468	US-07-951-715A-2	Sequence 2, Appl
21	32.2	1.5	3468	US-08-459-448A-2	Sequence 2, Appl
22	32.2	1.5	3468	US-08-459-595A-2	Sequence 2, Appl
23	32.2	1.5	3468	US-08-459-504B-2	Sequence 2, Appl
24	32.2	1.5	3468	US-08-459-444-2	Sequence 2, Appl
25	32.2	1.5	3468	US-09-053-549-3	Sequence 3, Appl
26	32.2	1.5	3468	US-09-547-422-2	Sequence 2, Appl
27	32	1.5	289	US-09-007-005-17	Sequence 17, Appl

28	32	1.5	289	US-09-244-796-17	Sequence 17, Appl
29	32	1.5	823	US-08-998-416-551	Sequence 551, App
30	32	1.5	1947	US-07-951-715A-3	Sequence 3, Appl
31	32	1.5	1947	US-08-459-448A-3	Sequence 3, Appl
32	32	1.5	1947	US-08-459-595A-3	Sequence 3, Appl
33	32	1.5	1947	US-08-459-504B-3	Sequence 3, Appl
34	32	1.5	1947	US-08-459-444-3	Sequence 3, Appl
35	32	1.5	1947	US-09-547-422-3	Sequence 3, Appl
36	32	1.5	2845	US-08-289-653-1	Sequence 1, Appl
37	32	1.5	3468	US-07-951-715A-4	Sequence 4, Appl
38	32	1.5	3468	US-07-951-715A-8	Sequence 8, Appl
39	32	1.5	3468	US-08-459-448A-4	Sequence 4, Appl
40	32	1.5	3468	US-08-459-595A-4	Sequence 4, Appl
41	32	1.5	3468	US-08-459-504B-4	Sequence 4, Appl
42	32	1.5	3468	US-08-459-595A-8	Sequence 8, Appl
43	32	1.5	3468	US-08-459-504B-4	Sequence 4, Appl
44	32	1.5	3468	US-08-459-504B-8	Sequence 8, Appl
45	32	1.5	3468	US-08-459-444-4	Sequence 4, Appl

ALIGNMENTS

```
RESULT 1
US-09-289-349-7
; Sequence 7, Application US/09289349
; Patent No. 6277574
; GENERAL INFORMATION:
; APPLICANT: Walker, Michael, G.
; APPLICANT: Volkmut, Wayne
; APPLICANT: Klingert, Tod, M.
; APPLICANT: Azimzai, Yalda
; APPLICANT: Yue, Henry
; TITLE OF INVENTION: GENES ASSOCIATED WITH DISEASES OF THE KIDNEY
; FILE REFERENCE: PB-0010 US
; CURRENT APPLICATION NUMBER: US/09/289,349
; CURRENT FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PERL Program
; SEQ ID NO 7
; LENGTH: 1090
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: 2673009CB1
US-09-289-349-7

Query Match      1.7%  Score 36.4;  DB 4;  Length 1090;
Best Local Similarity 49.5%;  Pred. No. 0.18;
Matches 94;  Conservative 0;  Mismatches 96;  Indels 0;  Gaps 0;

QY 1674 CCCCAGACGTGCTCTCCGAGGATGCCAATTCACATGCGTATGAGTACGCTCCAT 1733
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 136 CTTGACCGTCTTCTTACACCATCAAGCTCATGCACGACGACGACGACGATGATGCGT 195

QY 1734 CAAGGATATGAACCCAGAGAGCGTCCCTGCTACATCATACACAGGAGGCGCT 1793
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 196 CAGGACCAAGCATGCCCATTTGGGGCTTCCTCCCAAGAAATGACAGTATGGAGGC 255

QY 1794 ATTGAGGAAGCATGCTTAAGAGCATGATGTTGGCGTACAGAGAAATATCCATCCAC 1853
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 256 CTTGACCGTCTGATGATGGCGTGTGATGATGATGATGATGATGATGATGATGATGAT 315

QY 1854 CTTGACCGTCTGATGATGGCGTGTGATGATGATGATGATGATGATGATGATGATGAT 315
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 316 CTTGACCGTCTGATGATGGCGTGTGATGATGATGATGATGATGATGATGATGATGAT 315

RESULT 2
US-08-415-818-5
; Sequence 5, Application US/08415818
; Patent No. 5621079
; GENERAL INFORMATION:
```

```

1  APPLICANT: Cascieri, Margaret A.
2  APPLICANT: Linemeyer, David L.
3  APPLICANT: Macnell, Douglas J.
4  APPLICANT: Shiao, Lin-Lin
5  APPLICANT: Strader, Catherine D.
6  APPLICANT: Tan, Carina P.
7  APPLICANT: Weinberg, David H.
8  TITLE OF INVENTION: NEUROPEPTIDE Y RECEPTOR
9  NUMBER OF SEQUENCES: 12
10 CORRESPONDENCE ADDRESS:
11 ADDRESSEE: Mary A. Appollina
12 STREET: P.O. Box 2000, 126 E. Lincoln Ave.
13 CITY: Rahway
14 STATE: NJ
15 COUNTRY: USA
16 ZIP: 07065
17
18 COMPUTER READABLE FORM:
19 MEDIUM TYPE: Floppy disk
20 COMPUTER: IBM PC compatible
21 OPERATING SYSTEM: PC-DOS/MS-DOS
22 SOFTWARE: PatentIn Release #1.0, Version #1.30
23 CURRENT APPLICATION DATA:
24 APPLICATION NUMBER: US/08/415,818
25 FILING DATE:
26 CLASSIFICATION: 435
27 PRIOR APPLICATION DATA:
28 APPLICATION NUMBER: 08/383,746
29 FILING DATE: 03-FEB-1995
30 ATTORNEY/AGENT INFORMATION:
31 NAME: Appollina, Mary A.
32 REGISTRATION NUMBER: 34,087
33 REFERENCE/DOCKET NUMBER: 19390
34 TELECOMMUNICATION INFORMATION:
35 TELEPHONE: 908-594-3462
36 TELEFAX: 908-594-4720
37 INFORMATION FOR SEQ ID NO: 5:
38 SEQUENCE CHARACTERISTICS:
39 LENGTH: 2280 base pairs
40 TYPE: nucleic acid
41 STRANDEDNESS: double
42 TOPOLOGY: linear
43 MOLECULE TYPE: DNA (genomic)
44 HYPOTHETICAL: NO
45 ANTI-SENSE: NO
46 FEATURE:
47 NAME/KEY: CDS
48 LOCATION: 822..1937
49 US-08-415-818-5

```

```

Query Match 1.7%; Score 36.2; DB 1, Length 2280;
Best Local Similarity 5.0%; Pred. No. 0.35;
Matches 71; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 409 GACAAGTAGAAGACAGACCTCAAGGGTATCACTGACGAGATGGAGCGTTCGGCAAAGCT 468
      ||||| || ||||| ||||| ||||| |||||
Db 501 GACAAGGAGATCAGAAAGCTTTAGACATCATTTGTACATAGTGAGTTTGAGGAAAGCT 560
      ||||| ||||| ||||| ||||| ||||| |||||

QY 469 CAGAGGATATGCCAAGAACTGGCCAAAAAATGCAAGACAGAAACCAACTCTGCATC 528
      |||| |||| || |||| |||| |||| |||| ||||
Db 561 GAGGTATACATGAACCTCTCTCTCAAAAAACAAACAAACAAACCTTTTAC 620
      |||| |||| || |||| |||| |||| |||| ||||

QY 529 GTAGAGCCT 537
      || || ||
Db 621 TAATATCT 629

RESULT 3
US-08-894-236-5
: Sequence 5, Application US/08894236
: Patent No. 5939263
: GENERAL INFORMATION:
: APPLICANT: Casclerl, Margaret A.
: APPLICANT: Linemeyer, David L.

```

```

1  APPLICANT:  Macneil, Douglas J.
2  APPLICANT:  Shiao, Lin-Lin
3  APPLICANT:  Strader, Catherine D.
4  APPLICANT:  Tan, Carina P.
5  APPLICANT:  Weinberg, David H.
6  TITLE OF INVENTION:  NEUROPEPTIDE Y RECEPTOR
7  NUMBER OF SEQUENCES:  14
8  CORRESPONDENCE ADDRESS:
9  ADDRESSEE:  Mary A. Appollina
10 STREET:  P.O. Box 2000, 126 E. Lincoln Ave.
11 CITY:  Rahway
12 STATE:  NJ
13 COUNTRY:  USA
14 ZIP:  07065
15 COMPUTER READABLE FORM:
16 MEDIUM TYPE:  Floppy disk
17 COMPUTER:  IBM PC compatible
18 OPERATING SYSTEM:  PC-DOS/MS-DOS
19 SOFTWARE:  PatentIn Release #1.0, Version #1.30
20 CURRENT APPLICATION DATA:
21 APPLICATION NUMBER:  US/08/894,236
22 FILING DATE:
23 CLASSIFICATION:  435
24 PRIOR APPLICATION DATA:
25 APPLICATION NUMBER:  08/283,746
26 FILING DATE:  03-FEB-1995
27 PRIOR APPLICATION DATA:
28 APPLICATION NUMBER:  08/415,818
29 FILING DATE:  03-APR-1995
30 ATTORNEY/AGENT INFORMATION:
31 NAME:  Appollina, Mary A.
32 REGISTRATION NUMBER:  34,087
33 REFERENCE/DOCKET NUMBER:  19390Y
34 TELECOMMUNICATION INFORMATION:
35 TELEPHONE:  908-594-3462
36 TELEFAX:  908-594-4720
37 INFORMATION FOR SEQ ID NO:  5:
38 SEQUENCE CHARACTERISTICS:
39 LENGTH:  2280 base pairs
40 TYPE:  nucleic acid
41 STRANDEDNESS:  double
42 TOPOLOGY:  linear
43 MOLECULE TYPE:  DNA (genomic)
44 HYPOTHETICAL:  NO
45 ANTI-SENSE:  NO
46 FEATURE:
47 NAME/KEY:  CDS
48 LOCATION:  822..1937
49 US-08-894-236-5

```

```

Query Match      1.7%; Score 36.2; DB 2, Length 2280;
Best Local Similarity 55.0%; Pred. NO. 0.35;
Matches 71; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY 409 GACAAGTAGAAGGACAGCTCAAGGATATCACTGACGAGATGACGCTCTCGCGAAAGCT 468
      |||||  ||  ||||| ||||| ||||| ||||| ||||| |||||
Db 501 GACAAGGAGAGATCAAAAGCTTAAGACATCATTTGTACATAGTGAGTTTGAGAAAGCT 560
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 469 CAGGAGGATAGCCAAAGAACTGGCCAAAAAAGAAAATGCAGACGAGAACCACTCTGCATC 528
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 561 GAGGTTACATGSAACTCTCTCTCTCTCAAAAAACAAAACAAAACCAAAACCTTTAC 620
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 529 GTAGAGCCT 537
      |||||
Db 621 TAATATCT 629

```

```

1 APPLICANT: MacNeil, Douglas J.
2 APPLICANT: Shiao, Lin-Lin
3 APPLICANT: Strader, Catherine D.
4 APPLICANT: Tan, Carina P.
5 APPLICANT: Weinberg, David H.
6 TITLE OF INVENTION: NEUROPEPTIDE Y RECEPTOR
7 NUMBER OF SEQUENCES: 14
8 CORRESPONDENCE ADDRESS:
9 ADDRESSEE: Mary A. Appollina
10 STREET: P.O. Box 2000, 126 E. Lincoln Ave.
11 CITY: Rahway
12 STATE: NJ
13 COUNTRY: USA
14 ZIP: 07065
15
16 COMPUTER READABLE FORM:
17 MEDIUM TYPE: Floppy disk
18 COMPUTER: IBM PC compatible
19 OPERATING SYSTEM: PC-DOS/MS-DOS
20 SOFTWARE: Patentln Release #1.0, Version #1.30
21 CURRENT APPLICATION DATA:
22 APPLICATION NUMBER: PCT/US96/01444
23 FILING DATE:
24
25 CLASSIFICATION:
26 PRIOR APPLICATION DATA:
27 APPLICATION NUMBER: 08/383,746
28 FILING DATE: 03-FEB-1995
29 PRIOR APPLICATION DATA:
30 APPLICATION NUMBER: 08/415,818
31 FILING DATE: 03-APR-1995
32 ATTORNEY/AGENT INFORMATION:
33 NAME: Appollina, Mary A.
34 REGISTRATION NUMBER: 34,087
35 REFERENCE/DOCKET NUMBER: 19390Y
36 TELECOMMUNICATION INFORMATION:
37 TELEPHONE: 908-594-3462
38 TELEFAX: 908-594-4720
39
40 INFORMATION FOR SEQ ID NO: 5:
41
42 SEQUENCE CHARACTERISTICS:
43 LENGTH: 2280 base pairs
44 TYPE: nucleic acid
45 STRANDEDNESS: double
46 TOPOLOGY: linear
47 MOLECULE TYPE: DNA (genomic)
48 HYPOTHEetical: NO
49 ANTI-SENSE: NO
50 FEATURE:
51 NAME/KEY: CDS
52 LOCATION: 822..1937
53 PCT-US96-01444-5
54
55 Query Match 1.7%; Score 36.2; DB 5; Length 2280;
56 Best Local Similarity 55.0%; Pred. No. 0.35;
57 Matches 71; Conservative 0; Mismatches 58; Indels 0; Gaps 0;
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99

```

[illegible]

CITY: BOSTON
STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/004,139B
FILING DATE: 09-DEC-1992
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: RESNICK, DAVID S.
REGISTRATION NUMBER: 34235
REFERENCE/DOCKET NUMBER: 42188
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 523-3400
TELEFAX: (617) 523-6440
TELEX: 200291 STRE UR
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5837 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: not relevant
US-08-004-139B-1

Query Match 1.6%; Score 34; DB 1; Length 5837;
Best Local Similarity 57.5%; Pred. No. 3.2;
Matches 61; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 1405 GTCATGCAAGAAATTCAGAGCGACAGAGAGATATGACACTTCGATTGACAG 1464
DB 3686 GTCATGTCACAGATAGTCTTCAACACCTTCGAGAAATGCAATATCTTTTGGCAAT 3627
QY 1465 AGTGTGTCCTTATAGCGACAAGTTCATGCCATCTCAAGTCCA 1510
DB 3626 GCTGTAGTCACCTTAGAGAAAAGATCTTTATTTTCACAAATCTA 3581

RESULT 9
US-08-117-491-1/c
Sequence 1, Application US/08117491
Patent No. 5500363
GENERAL INFORMATION:
APPLICANT: Comb, Donald G.
APPLICANT: Perler, Francine
APPLICANT: Kucera, Rebecca
APPLICANT: Jack, William F.
TITLE OF INVENTION: RECOMBINANT THERMOSTABLE DNA
TITLE OF INVENTION: POLYMERS FROM ARCHAEABACTERIA
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS
STREET: 130 WATER STREET
CITY: BOSTON
STATE: MASSACHUSETTS
COUNTRY: US
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/117,491
FILING DATE: 07-SEP-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/811,421

FILING DATE: 12-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/686,340
FILING DATE: 17-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/626,057
FILING DATE: 11-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/513,994
FILING DATE: 26-APR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Resnick, David S.
REGISTRATION NUMBER: 34,235
REFERENCE/DOCKET NUMBER: 39296C3FWC2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 523-3400
TELEFAX: (617) 523-6440
TELEX: 200291 STRE UR
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5837 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-117-491-1

Query Match 1.6%; Score 34; DB 1; Length 5837;
Best Local Similarity 57.5%; Pred. No. 3.2;
Matches 61; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 1405 GTCATGCAAGAAATTCAGAGCGACAGAGAGATATGACACTTCGATTGACAG 1464
DB 3686 GTCATGTCACAGATAGTCTTCAACACCTTCGAGAAATGCAATATCTTTTGGCAAT 3627
QY 1465 AGTGTGTCCTTATAGCGACAAGTTCATGCCATCTCAAGTCCA 1510
DB 3626 GCTGTAGTCACCTTAGAGAAAAGATCTTTATTTTCACAAATCTA 3581

RESULT 10
US-08-271-364A-1/c
Sequence 1, Application US/08271364A
Patent No. 5756334
GENERAL INFORMATION:
APPLICANT: PERLER, FRANCINE B.
APPLICANT: SOUTHWORTH, MAURICE W.
TITLE OF INVENTION: RECOMBINANT THERMOSTABLE DNA POLYMERASE
TITLE OF INVENTION: FROM ARCHAEABACTERIA
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: NEW ENGLAND BIOLABS, INC.
STREET: 32 TOZER ROAD
CITY: BEVERLY
STATE: MASSACHUSETTS
COUNTRY: US
ZIP: 01915
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/271,364A
FILING DATE: 06-JUL-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/811,421
FILING DATE: 18-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/686,340
FILING DATE: 17-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/626,057

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/626,057
FILING DATE: 11-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/513,994
FILING DATE: 26-APR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Williams, Gregory D.
REGISTRATION NUMBER: 30901
REFERENCE/DOCKET NUMBER: NEB-054C3FC2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (508) 927-5054
TELEFAX: (508) 927-1705
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5837 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-222-715B-1

Query Match
Best Local Similarity 1.6%; Score 34; DB 2; Length 5837;
Matches 61; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 1405 GTAATCGACAGAAATTCAGGACGACGAGAGAGTATGCACTTCTATTCAGCAAG 1464
DB 3686 GTCATCGTCACAGTAGTGTCTTCAACACCTTCGAGAAATGCAATCTTTTCCGCAAT 3627

QY 1465 AGTGTGTTCTTATAGCGACAGTTCATGCCATGCTCAAGTCCA 1510
DB 3626 GCTGTAGTCCACCTTAGAGAAAAGATCCTTTATTTTCACAAATCTA 3581

RESULT 13
PCT-US96-10545A-1/c
Sequence 1, Application PC/TUS9610545A
GENERAL INFORMATION:
APPLICANT: COMB, DONALD G.
APPLICANT: PERLER, FRANCINE B.
APPLICANT: JACK, WILLIAM E.
APPLICANT: XU, MING-QUN
APPLICANT: HODGES, ROBERT A.
APPLICANT: NOREN, CHRISTOPHER J.
TITLE OF INVENTION: MODIFIED PROTEINS AND METHODS OF THEIR
TITLE OF INVENTION: PRODUCTION
NUMBER OF SEQUENCES: 77
CORRESPONDENCE ADDRESS:
ADDRESSEE: GREGORY D. WILLIAMS, NEW ENGLAND BIOLOGICALS, INC.
STREET: 32 TOWER ROAD
CITY: BEVERLY
STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 01915
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/10545A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/580,555
FILING DATE: 29-DEC-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/496,247
FILING DATE: 28-JUN-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/146,885
FILING DATE: 03-NOV-1993
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/004,139
FILING DATE: 09-DEC-1992
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: WILLIAMS, GREGORY D.
REGISTRATION NUMBER: 30901
REFERENCE/DOCKET NUMBER: NEB-036C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (508) 927-5054
TELEFAX: (508) 927-1705
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5837 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: not relevant
PCT-US96-10545A-1

Query Match
Best Local Similarity 1.6%; Score 34; DB 5; Length 5837;
Matches 61; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 1405 GTAATCGACAGAAATTCAGGACGACGAGAGAGTATGCACTTCTATTCAGCAAG 1464
DB 3686 GTCATCGTCACAGTAGTGTCTTCAACACCTTCGAGAAATGCAATCTTTTCCGCAAT 3627

QY 1465 AGTGTGTTCTTATAGCGACAGTTCATGCCATGCTCAAGTCCA 1510
DB 3626 GCTGTAGTCCACCTTAGAGAAAAGATCCTTTATTTTCACAAATCTA 3581

RESULT 14
US-09-773-816-1
Sequence 1, Application US/09773816
Patent No. 6340774
GENERAL INFORMATION:
APPLICANT: Stanford University
APPLICANT: Khosla, Chaitan
TITLE OF INVENTION: NON-SPEROIDAL ESTROGEN-RECEPTOR
TITLE OF INVENTION: ANTAGONISTS
FILE REFERENCE: 28600-20210.00
CURRENT APPLICATION NUMBER: US/09/773,816
CURRENT FILING DATE: 2001-06-08
PRIOR APPLICATION NUMBER: US 60/243,458
PRIOR FILING DATE: 2000-10-25
PRIOR APPLICATION NUMBER: US 60/179,305
PRIOR FILING DATE: 2000-01-31
NUMBER OF SEQ ID NOS: 1
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 23673
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc.feature
LOCATION: (1)-(23623)
OTHER INFORMATION: n = A,T,C or G
US-09-773-816-1

Query Match
Best Local Similarity 1.6%; Score 33.6; DB 4; Length 23673;
Matches 78; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 625 AAGTTCGAGCGGATACGACACATGATGTGCGCGCTCACACGCGCACTTCAGCGTA 684
DB 12747 AAGTTCCTTCAGCGGACGACCGTGTGCGGCGGTGTTCCTCCGTCGCGCGCGCGTC 12806

QY 685 TTCCGCGGTATAGCGGTCGACGACGCGCGCGGCGGGAATACAGCAGAGACATAAACCC 744

Db 12807 TTCCGGGTGCTGTACGCGTTGACCGCGCGGCGGTGCGGGCGGACGCGCCCTCC 12866

OY 745 TATAAGCCCGTTTACTTGTGCTGCCGTATCCAT 776

Db 12867 TTCCGAGGAGTCAACGACGCGCGCCGACCAT 12898

RESULT 15

US-09-428-517-1
; Sequence 1, Application US/09428517
; Patent No. 6251636
; GENERAL INFORMATION:
; APPLICANT: Bellach, Mary C.
; APPLICANT: Shah, Sanjay Krishnakant
; APPLICANT: Mcdaniel, Robert
; APPLICANT: Tang, Li
; TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 30062-20029.00
; CURRENT APPLICATION NUMBER: US/09/428,517
; CURRENT FILING DATE: 1999-10-28
; EARLIER APPLICATION NUMBER: 60/120,254
; EARLIER FILING DATE: 1999-02-16
; EARLIER APPLICATION NUMBER: 60/106,100
; EARLIER FILING DATE: 1998-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 50937
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Recombinant DNA
US-09-428-517-1

Query Match 1.6%; Score 33.6; DB 4; Length 50937;
Best Local Similarity 47.2%; Pred. No. 18;
Matches 102; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

OY 1377 CGACAAGCTCCCGATATATCAAGATGTATGCAAGAAATCAAGCGACAGAA 1436
Db 44477 CGCCGAGATCTACGACCGCGGTACCGGGCCGGGCAAGACTACGCCGGCGAGCGAA 44536
OY 1437 GAAGTATCAGACTTCTATTTCGACACAGAGTGTGCTCTTATAGCACAAGTTCATGC 1496
Db 44537 GGACGTGGCGGACCTCGTGGCGGACCGGGTGCGGACGCGTCTCTCGGACGTGGC 44596
OY 1497 CATGCTCAAGTCATGACAAAGAAAGTTTCCCAAGGCTATCGAANAAGATCCGGCAGT 1556
Db 44597 CTGCGGCACGGGCGGCGACCTGCGGCACTTGCACGCTCTTCGACGAGCGCCGGGTCT 44656
OY 1557 AGAGCTTCCAAAGAGGCTAATAGCTGCTGCGGCGC 1592
Db 44657 CGAAGTGTCCGCGACATGCTGGACATGCGCCGCTC 44692

Search completed: October 17, 2002, 21:41:49
Job time : 243 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 17, 2002, 19:23:15 : Search time 1728 Seconds

(without alignments)
16707.168 Million cell updates/sec

Title: US-10-008-355-1
Perfect score: 2139
Sequence: 1 atgcaatgaaatgaaag.....aagagctgagttgctctaa 2139

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 674847542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_estl:*
10: gb_estc2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pin:*
16: em_gss_vrl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	110	5.1	537	12	BH388186 AG-ND-124
C 2	110	5.1	591	12	BH392840 AG-ND-162
C 3	108.2	5.1	500	12	BH386991 AG-ND-148
C 4	103.8	4.9	418	12	BH375201 AG-ND-173
C 5	72.4	3.4	426	12	BH400391 AG-ND-147
C 6	66.4	3.1	697	12	BH375688 AG-ND-120
C 7	55	2.6	398	9	AW774807 EST333958
C 8	49.2	2.3	1005	12	BH164253 ENT7637F
C 9	42.8	2.0	555	12	BH400866 AG-ND-125
C 10	42.2	2.0	692	9	A1405287 GH25242.5
C 11	39.4	1.9	721	12	BH391233 AG-ND-141
C 12	39.4	1.8	822	12	BH371846 AG-ND-162
C 13	38.6	1.8	687	12	BH448937 BOC2L67R
C 14	38.4	1.8	184	9	AW375157 OV2-CT012
C 15	38.4	1.8	466	10	BF426788 d172a12.y
C 16	38.4	1.8	491	9	AW148212 da13h05.x
C 17	38.4	1.8	895	10	BF784434 602110820

18	38	1.8	465	9	AV628406
19	38	1.8	497	9	AV621707
20	38	1.8	587	10	B1247683
C 21	38	1.8	1091	10	BE282103
C 22	37.8	1.8	422	9	BB851067
C 23	37.8	1.8	461	9	AL372451
C 24	37.8	1.8	680	10	B1309279
C 25	37.8	1.8	773	10	BE644863
C 26	37.6	1.8	422	9	AV629946
27	37.6	1.8	927	10	BC437679
28	37.4	1.7	367	10	BG983576
29	37.2	1.7	318	9	AV644201
30	37.2	1.7	348	10	BE717547
31	36.8	1.7	668	10	AU070770
32	36.6	1.7	313	10	BE804382
33	36.6	1.7	436	9	AV637236
34	36.6	1.7	471	9	AV640279
35	36.6	1.7	478	9	AV643425
36	36.6	1.7	497	10	BE552689
37	36.6	1.7	509	10	BE252639
38	36.6	1.7	517	9	AV388690
39	36.6	1.7	517	9	AV635803
40	36.6	1.7	552	10	B1643498
41	36.6	1.7	586	10	B1675086
42	36.4	1.7	409	9	AA277303
43	36.4	1.7	459	9	AW154047
44	36.4	1.7	482	9	AA541836
C 45	36.4	1.7	515	9	AV743237

ALIGNMENTS

RESULT 1
BH388186/c
LOCUS
DEFINITION AG-ND-124P3.TF ND-TAM Anopheles gambiae genomic clone AG-ND-124P3,
DNA sequence.
ACCESSION BH388186
VERSION BH388186.1 GI:17334327
KEYWORDS GSS.
SOURCE African malaria mosquito.
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae
; Anophelinae.
REFERENCE
AUTHORS Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.
TITLE Direct Submission of BAC-end sequences from Anopheles gambiae
JOURNAL Unpublished (2001)
COMMENT
Other_GSSs: AG-ND-124P3.TF
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: b.loftus@tigr.org
This clone is from an A. gambiae BAC library (ND-TAM) provided by
F.H. Collins and sequenced by The Institute for Genomic Research
(TIGR). The BAC library was generated from A. gambiae PEST strain
DNA. All DNA was extracted from newly hatched first instar larvae
to minimize the inclusion of DNA from microorganisms that inhabit
the gut. The DNA is derived from mixed sexes of larvae. The BAC
library was constructed at Texas A&M University BAC Center
University, College Station, Texas 77843-2123, USA using a HindIII
partial digest.
Seq primer: M13 For
Class: BAC ends.
Location/Qualifiers
1..537
/organism="Anopheles gambiae"
/strain="PEST"

FEATURES
source

F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from *A. gambiae* PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.

Seq primer: M13 Rev

Class: BAC ends.

Location/Qualifiers

1. 500

/organism="Anopheles gambiae"

/strain="PEST"

/db_xref="taxon:7165"

/clone="AG-ND-148L11"

/clone.lib="ND-TAM"

/note="Vector: pECBAC1; Site_1: HindIII"

BASE COUNT

143 a 105 c 90 g 162 t

ORIGIN

Query Match

Best Local Similarity 57.6%; Score 108.2; DB 12; Length 500;

Matches 194; Conservative 0; Mismatches 143; Indels 0; Gaps 0;

527 TCGAGAGCCTTCTATTCACAGCAAGATCTCCATCGTCTACGATGATTCAGG 586

369 TGGTGAAGATTTCTTCAAGAAATGATTTACTTGTATTCAGAGATTTCAAG 310

587 ACCTGCTATGATATTTGCTCTCCAGCTCTAGTAGTTCGAGGCGATACGACA 646

309 ATGTGCTGTAGAGGATCTCCGCCATCTATTTGTAATATGCGATGATGATATA 250

647 ACTGATGTGGCCGCTCAGACGCGGCACTTACAGCTATTCGCGGTATGCGGTCGG 706

249 ACTGGGAATGGCCAGACACAGAGACTTCTGTTTCCGCTTATGCTGATMAAA 150

707 ACAACGGCGCGCGAATACAGCAAGCAATAAACCTTAAGCCGTTACTTGCTG 766

189 ATGGCAATCTCTGCTGATATTCAGCTACACAGCTTCTTAAGCCTTAACATCATTTAC 130

767 CCGTATCCATGCAAGGCTACAGGCTGACGACTATGACCATGACCATGCTTCCCGGCA 826

129 CAATTCACATAAAGGAATTAAGCTCGTGATTTTGAATGATGATGATATCCGTGTA 70

827 GTACGATCGCTACCTCCTCTTGGGCTGGAAGA 863

69 CAACAAACAGATACCTTACTTCTTGGATCGAACA 33

DB BH375201 418 bp DNA linear GSS 10-DEC-2001

AG-ND-173C3.TR ND-TAM Anopheles gambiae genomic clone AG-ND-173C3,

DEFINITION DNA sequence.

ACCESSION BH375201

VERSION BH375201.1 GI:17321343

KEYWORDS GSS.

SOURCE African malaria mosquito.

ORGANISM Anopheles gambiae

REFERENCE 1 (bases 1 to 418)

AUTHORS Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.

TITLE Direct Submission of BAC-end sequences from Anopheles gambiae

JOURNAL Unpublished (2001)

COMMENT Contact: Brendan J Loftus

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0208

Fax: 301 838 3543

Email: bjlloftus@tigr.org

This clone is from an *A. gambiae* BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from *A. gambiae* PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.

Seq primer: M13 Rev

Class: BAC ends.

Location/Qualifiers

1. 418

/organism="Anopheles gambiae"

/strain="PEST"

/db_xref="taxon:7165"

/clone="AG-ND-173C3"

/clone.lib="ND-TAM"

/note="Vector: pECBAC1; Site_1: HindIII"

BASE COUNT

145 a 81 c 70 g 122 t

ORIGIN

Query Match

Best Local Similarity 54.6%; Score 103.8; DB 12; Length 418;

Matches 207; Conservative 0; Mismatches 172; Indels 0; Gaps 0;

530 TAGAGCCTTCTATTCACAGCAAGATCTCCATCGTCTACGATGATTCAGGACG 589

30 TAAAGTCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 89

590 TTGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 649

90 TCCGTTTATGAGAGGCTCCACCTCAATCTATTTGTTGTTGTTGTTGTTGTTGTT 149

650 GGATGTGGCGCGCTCAGACGCGGAGCTTACGCTATTCGCGGTATGCGGCGGACA 709

150 GGGTATGCGCAACACATACAGAGAGACTCTCTATGTTCCGTATCTATGACAGAAAACA 209

710 ACCGCGCGCGCGATACAGCAAGCAATTAACCCCTTAAGCCGTTACTTGCTGCGG 769

210 ATTAACCTCGAGATATTTCTTAAGATTAATCCATTCATCCGAAATATTCCTCCG 269

770 TATCCATGCAAGGCTACAGGCTGACGACATGACATGACATGACATGACATGACAT 829

270 TATCTGTAAGAAAGCCTTAAGAAAGTGATTTCAATTCGTTTGGGATTCGCGGAAAA 329

830 CGGATCGCTACCTCCTCTTGGGCTGTGGAAGATGATGCAAAAGAGAACATCTTC 889

330 CAACCTGATACCTTCTCTCAATTTGCTGTAAGAAAGATCAATCAATCAATCAATCA 389

890 GTATCGAATTCGCGGAT 908

390 AAATTAACGATCGTATAT 408

DB BH400391/C 426 bp DNA linear GSS 11-DEC-2001

AG-ND-147H4.TF ND-TAM Anopheles gambiae genomic clone AG-ND-147H4,

DEFINITION DNA sequence.

ACCESSION BH400391

VERSION BH400391.1 GI:17346607

KEYWORDS GSS.

SOURCE African malaria mosquito.

ORGANISM Anopheles gambiae

REFERENCE 1 (bases 1 to 426)

AUTHORS Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.

TITLE Direct Submission of BAC-end sequences from Anopheles gambiae

JOURNAL Unpublished (2001)

COMMENT
Other_GSSs: AG-ND-147H4.TR
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjoftus@ligr.org
This clone is from an A. gambiae BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.
Seq primer: M13 For
Class: BAC ends.
Location/Qualifiers
source 1..426
/organism="Anopheles gambiae"
/strain="PEST"
/db_xref="taxon:7165"
/clone="AG-ND-147H4"
/clone_11b="ND-TAM"
/note="Vector: pECBAC1; Site_1: HindIII"

BASE COUNT 111 a 79 c 76 g 160 t
ORIGIN

Query Match 3.4%; Score 72.4; DB 12; Length 426;
Best Local Similarity 57.8%; Pred. No. 3.6e-09;
Matches 149; Conservative 0; Mismatches 106; Indels 3; Gaps 1;

QY 1756 GGTGCTGTGACACATATATACGAGGCAAGGCGCTATTGGAGACGATCCTAAG 1815
|||||
DB 258 GGTATACACAGATATACATACCAATGAGGATATATTAAGAAAGT 199
QY 1816 AGCATGATTTGGCCGTACAGAGAAATATCTCGACCTTCCGACCAAAATATGGT 1875
|||||
DB 198 GACGAAGATTCGATCTTCACAGACCTCTTGATCTTTATTAAGAAAAATTTGGT 139
QY 1876 ---CGCTATGCGAGAGCGTCATCTCATATCGCTTCTTATCAAGACGATCAG 1932
|||||
DB 138 ATGTACAAAGACAAAGACGGCACTTCATGTAACCTCTTCTTAATACGATATACA 79
QY 1933 GGGGTAATCTCCGATAGCCCGGATTCGATAGAGCGCCGCTGATCGCTTTC 1992
|||||
DB 78 GGAGTAACCTCAAGCTTCTCAATATATGATGTTACGGAAGACTTATAGTCTTGATTT 19
QY 1993 GATGCACTGGGAGCT 2010
|||
DB 18 GACGGAAACAGTGAAGCT 1

RESULT 6
BH375688/c 697 bp DNA linear GSS 10-DEC-2001
LOCUS AG-ND-120J9.TR ND-TAM Anopheles gambiae genomic clone AG-ND-120J9,
DEFINITION DNA sequence.
ACCESSION BH375688
VERSION BH375688.1 GI:17321830
KEYWORDS GSS.
SOURCE African malaria mosquito.
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae
; Anopheles.
REFERENCE 1 (bases 1 to 697)
AUTHORS Shetty J., Malek J., Koo H., Collins F., Gardner M. and Loftus B.J.
TITLE Direct Submission of BAC-end sequences from Anopheles gambiae
JOURNAL Unpublished (2001)
COMMENT Other_GSSs: AG-ND-120J9.TR

Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjoftus@ligr.org
This clone is from an A. gambiae BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.
Seq primer: M13 Rev
Class: BAC ends.
Location/Qualifiers
source 1..697
/organism="Anopheles gambiae"
/strain="PEST"
/db_xref="taxon:7165"
/clone="AG-ND-120J9"
/clone_11b="ND-TAM"
/note="Vector: pECBAC1; Site_1: HindIII"

BASE COUNT 214 a 133 c 122 g 228 t
ORIGIN

Query Match 3.1%; Score 66.4; DB 12; Length 697;
Best Local Similarity 50.3%; Pred. No. 2.2e-07;
Matches 197; Conservative 0; Mismatches 186; Indels 9; Gaps 1;

QY 16 AAAGATATCTCTCGGAGCAGCCCGCTGTGGGCTTCAGGGGTAGCCAAAGCCGAC 75
|||||
DB 475 AAAAATCTCTATTAGCACTACCTTACTCCCGAGAGCGCTTTGCTCAACAGGCG 416
QY 76 AAAGCATGTGCGCTCCCTCAACGACATCATGAGAAATCTGATCGATCGGTAGCTC 135
|||||
DB 415 GGAGGATGTGATTCCTACAGAGATTAAATGAAGAA-----AATGAAGAAATTG 365
QY 136 GCGTTAGCGCTCCGTTGGATTCGCTACAGTTTCGACACGCGCTCCATTCGCAATGCC 195
|||||
DB 364 GCATGAGAGATTTCGCGAAGATATTTCAATACCTCAAAAACCTAGTATTAAGATGCT 305
QY 196 GTGGTTATCTTGGTGGCGGATGATCCGCTATCACAGTGCAGATCGAGCGCTGATCTT 255
|||||
DB 304 GTAGTGCATTTCAACGAGCGCTGACTGCTGAATTTCTCCAAAAGTTTGTTATTG 245
QY 256 ACCAACCACTGCGGATATGCTCTATCCAGAGCCAAAGCAGCGGTGATCAGACTAT 315
|||||
DB 244 ACAAAATCACCACTGTGCGCTTCGACAGATTCAGCTCANTGCATCAAAATATACCTT 185
QY 316 CTGCGGATGTTTGGTTCTTCGCACGATGGGTGAGAGACTTCGATTCGGGCTTTCC 375
|||||
DB 184 CTAAATGATGTTTCTGGGCAAAAAACATGGGAGAAACTTCCAAACCTGTGTGTT 125
QY 376 GTGAAGTATCTCGCAAGATCTGAAGTAAAC 407
|||
DB 124 GTAGATTTCATCACCAGATATTAAAGAGTGAC 93

RESULT 7
AW774807 398 bp mRNA linear EST 07-SEP-2000
LOCUS AW774807
DEFINITION EST333958 KV3 Medicago truncatula cDNA clone PKV3-24C3, mRNA
sequence.
ACCESSION AW774807
VERSION AW774807.1 GI:7718724
KEYWORDS EST.
SOURCE batrel medic.
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

REFERENCE	TITLE	FEATURES
1 (bases 1 to 398)	ESTs from roots of Medicago truncatula after Rhizobium inoculation	
Medicago.		
Vandenbosch, K., Hurt, J., Moore, J., Beremand, P., Peng, H., Ellis, L., Town, C.D., Bowman, C.L., Craven, M.B., Hansen, T.S., Holt, I.E. and Frazer, C.M.		
Unpublished (1999)		
Contact: Vandenbosch K		
Department of Biology		
Texas A&M University		
College Station, TX 77843-3258, USA		
Tel: 409 845 7707		
Fax: 409 845 2891		
Email: kate@mail.bio.tamu.edu		
Texas A&M Esp name: T258226e		
TIGR sequence name: MTEBG14TK		
More information is available at:		
http://chrhysie.tamu.edu/medicago		
Seq primer: Skmod (CTA gaa cta gta ggt gat cc).		
Location/Qualifiers		
1..398		
/organism="Medicago truncatula"		
/cultivar="genotype A17"		
/db_xref="taxon:3880"		
/clone="pKV3-24C3"		
/clone_1id="KV3"		
/divs_type="Seeding roots"		
/des_stage="3 days post-inoculation with Sinorhizobium meliloti"		
/lab_host="E. coli strain XL0LR"		
/note="Vector: plusscript SK -; Site_1: EcoRI; Site_2: XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA was directionally ligated into the Unizap XR vector from stragene and packaged using gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-Zap phage using Ex-assist helper phage and propagated in XL0LR cells."		
BASE COUNT	137 a	64 c 81 g 116 t
ORIGIN		
Query Match	2.6%;	Score 50; DB 9; Length 398;
Best Local Similarity	50.6%;	Pred. NO. 0.00026;
Matches 133; Conservative	0;	Mismatches 130; Indels 0; Gaps 0;
OY	67	AAAGCGCAAAAGCGATGTGGCTCTCAACGAATCAATCAGAGAAATCTGGATGAATG 126
Db	28	AGAGCGGATGAAGGAATGTGGTCTCGATGTTATCGAAAGATTAATCACAGGATATG 87
OY	127	CGTGGCTGGGCTTACGCTCCCGTGGATTCGCTCAAGTTTGCACAGCCGTCAT 186
Db	88	CAAAAATGGGGCTTCAATTGACACGCCAAGAAATTTACAGATTACAAATCACAGCTTA 147
OY	187	GCCATGCGCGTGTATTTGCGTGGCGGATACCGGTATTCACAGTGTCCGATAGAGC 246
Db	148	AAAGATGGGATTTGCAATTCATGATGGGGGTTGTATGCAAGAAATGTTTCAAAACGCGT 207
OY	247	CTGATCTTACCAACCACTGCGGATACGCTGCTATTCACAGCCAAAGCAGGTGAT 306
Db	208	TTGGTTCTTACCAACCACTGCTGTTATAGCGGATTCAGAACTTTCATCTCAGAA 267
OY	307	CACGACTATCGCGGATGGTTT 329
Db	268	CAAAATCAATTAATAAATGCTTT 290
RESULT 8		
BH164253/c	1005 bp	DNA linear GSS 24-SEP-2001
LOCUS	BH164253	
DEFINITION	ENTR6337F Entamoeba histolytica Sheared DNA Entamoeba histolytica genomic, DNA sequence.	
ACCESSION	BH164253	

VERSION	BH164253.1	GI:15737691
KEYWORDS	GSS.	
SOURCE	Entamoeba histolytica.	
ORGANISM	Entamoeba histolytica	
REFERENCE	Entamoebidae; Entamoeba.	
AUTHORS	1 (bases 1 to 1005)	
TITLE	Loftus, B., Wang, Z., Van Aken, S. and Fraser, C.	
JOURNAL	Determination of clone end sequences from Entamoeba histolytica	
COMMENT	HMI:IMSS sheared DNA library (2001)	
	Unpublished (2001)	
	Contact: Brendan J Loftus	
	Department of Eukaryotic Genomics	
	The Institute for Genomic Research	
	9712 Medical Center Dr., Rockville, MD 20850, USA	
	Tel: 301 838 0208	
	Fax: 301 838 3543	
	Email: b1loftus@tigr.org	
	Clones are derived from the Entamoeba histolytica HMI:IMSS sheared	
	DNA library	
	Seq primer: M13-Forward	
	Class: shotgun	
	High quality sequence start: 15	
	High quality sequence stop: 487.	
FEATURES	Location/Qualifiers	
Source	1..1005	
	/organism="Entamoeba histolytica"	
	/strain="HMI:IMSS"	
	/db_xref="taxon:5759"	
	/clone_lib="Entamoeba histolytica Sheared DNA"	
	/note="Vector: PHOSI; Site_1: Bst I; Constructed at The	
	Institute for Genomic Research (TIGR), Rockville, MD.	
	Genomic DNA isolated from broth cultures of E. histolytica	
	using a method described by Clark and Diamond (Clark,	
	C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a	
	method for isolate identification. Exp. Parasitol.	
	77:450.). The DNA was mechanically sheared to give a	
	tight size distribution (~2 kb). The v + i method used for	
	the library construction is described in detail in Smith,	
	H.O. and Venter, J.C. (Making small insert libraries for	
	whole genome shotgun sequencing projects. In Genome	
	Sequencing: A Practical Approach, eds. M. Vaundin and B.	
	Barell, Oxford University Press, 1999)."	
BASE COUNT	186 a 314 c 336 g 169 t	
ORIGIN		
Query Match	2.3%; Score 49.2; DB 12; Length 1005;	
Best Local Similarity	56.0%; Pred. No. 0.018;	
Matches	93; Conservative 0; Mismatches 73; Indels 0; Gaps 0;	
QY	195 CGTGGTATCTCTCGGATGGGATGACCGGGTATCAAGTGTCCGATCGAGGCGCTGATCCTT 254	
DB	229 GCCCGTGGGCGCGCGGGGATGACCGGGTCTTCGTTCCGTTCGCCGACGGGCTCTGCT 170	
QY	255 TACCAACCAACCACTCGGATACGGTGTATCCAGAGCCAAAGCAGCGTGATCAGACTA 314	
DB	169 GACCAACCAACCAACCGTGGGATGATTCAGTAAACAGCTCGCCGACGACGACT 110	
QY	315 TCTGGCGATGCTTTCGTTTCTCGACGATGGGTGAGAGCTCCG 360	
DB	109 GATCACTGATGCTTCATCGCCAAAGCGCGCTCGACGAGCGCCG 64	
RESULT 9		
LOCUS	BH400866/c	
DEFINITION	BH400866 555 bp DNA linear GSS 11-DEC-2001	
ACCESSION	AG-ND-125M4.TR ND-TAM Anopheles gambiae genomic clone AG-ND-125M4,	
VERSION	BH400866	
KEYWORDS	BH400866.1 GI:17347082	
SOURCE	GSS.	
ORGANISM	African malaria mosquito.	
	Anopheles gambiae	
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;	

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 555)	Koo, H., Collins, F., Gardner, M., and Loftus, B. J.	Direct Submission of BAC end sequences from Anopheles gambiae	Unpublished (2001)	
Other_GSSS: AG-ND-125M4_7F				
perlygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae				
Anopheles.				

pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoides; Anophelinae.

1 (bases 1 to 555)

Shetty, J., Malek, J., Koo, H., Collins, F., Gardner, M., and Loftus, B. J. Direct Submission of BAC-end Sequences from Anophelinae gambiae Unpublished (2001)

Other GSSs: AG-ND-125M4.TF

Contact: Brendan J Loftus

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 3208

Fax: 301 838 3208

This clone is from an *A. gambiae* BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from *A. gambiae* PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.

Seq primer: M13 Rev

class: BAC ends.

FEATURES	source	location/Qualifiers
	1..555	
		/organism="Anopheles gambiae"
		/strain="PEST"
		/db_xref="taxon:7165"
		/clone="AG-ND-125M4"
		/clone_1b="ND-TAM"
		/note="Vector: pCDBAC1; Site_1: HindIII"
BASE COUNT	152 a	105 c 94 g 204 t
ORIGIN		

Query Match	2.0%;	Score 42.8;	DB 12;	Length 555;
Best Local Similarity	66.0%;	Pred. No. 0.81;		
Matches 62;	Conservative 0;	Mismatches 32;	Indels 0;	Gaps 0;

OY	926	AGGAACCATATGCGCAGATAGCGTACCCTGATCAATTGCACAGAATATGCTAGA	985
Dδ	547	ATGAAATATATGCGTCATGATATATSCAACCCGTATTAATATGCACTCAATATATGCATCCG	488
OY	986	GTCCTAACAATTTGGAAAGATTTCCATCGGTATGAA	1019
Dδ	487	TGGCTAACTATTTGGAAAAAATATGTTGGTAGTA	454

RESULT 10	LOCUS	DEFINITION	LOCATION
AI405287	692 bp	mRNA	linear
AI405287	GH25242.5prlme	GH Drosophila melanogaster head pot2 Drosophila melanogaster cDNA clone GH25242.5prlme, mRNA sequence.	EST 19-APR-2001

VERSION	A1405287.1	GI:4248374
KEYWORDS	EST.	
SOURCE	fruit fly.	
ORGANISM	Drosophila	melanogaster

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Eukaryote; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha; Ephydroidea; drosophilidae; Drosophila.
1 (bases 1 to 692)
Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G., Lewis, S., and Rubin, G.M.: BDGP/HHMI Drosophila EST Project Unpublished (2001)
Contact: Stapleton, M.

Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: <http://www.fruitfly.org/EST>, esteff@lbl.gov

```

Plate: 252  row: D  column: 6
High quality sequence stop: 496.
location/Qualifiers
1. .692
FEATURES
source

```

```

/clone.lib="GH Drosophila melanogaster head port2"
/sex="male and female"
/dev.stage="adult"
/lab.host="DH5_+alpha"
/note="Organ: head; Vector: port2; Site_1: EcoRI; Site_2:
XhoI; Sized fractionated cDNAs were directly ligated into
port2. Plasmid cDNA library."

```

BASE COUNT	147 a	210 c	181 g	153 t	1 others
ORIGIN					

Query Match	2.0%;	Score 42.2;	DB 9;	Length 692;
Best Local Similarity	49.3%;	Pred. No. 1.3;		
Matches 110;	Conservative 0;	Mismatches 113;	Indels 0;	Gaps 0;

QY 1470 GGTTCCTTATAGCGACAAGTTCATGCCATGCTCAAGTCCATGGACAAAGAAAGTTTGC 1522

Db 254 GGGCCGCTATCAGCCGCCCTGCTGCTGCCCTCCCGAGAgtccAGCgactCGgttctgtttgt 313

1530 C A A G G C I A I C G A G A A A G A I C C G G C A G I A G A G C T I T C C A G A G C G I A A I A G C I G C I C G 1565

22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1041 1042 1043 1044 1045 1046 1047 1048 1049 1050 1051

[illegible]

Ov 1650 CCCCCCCCCCCCCCCCCCCCAACCTAATCCCGGAGCCTTGGTCCTGTCCGC
1692

Db 434 CATCATCCTGTGGTTCATAGTCTGCTTCGACTGCTCTGCTG 476

RESULT 11
BH391233

LOCUS BH391233 721 bp DNA linear GSS 11-DEC-2001
 DEFINITION AG-ND-141L2, "F ND-7AM Anopheles gambiae genomic clone AG-ND-141L2,
 DNA sequence."
 ACCESSION BH391233
 VERSION BH391233.1 GI:17337374
 KEYWORDS GSS.
 SOURCE African malaria mosquito.

ORGANISM	REFERENCE
Anopheles gambiae	1 (bases 1 to 721)
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoides; Anopheles.	Shetty, J., Malek, J., Koo, H., Collins, F., Gardner, M., and Loftus, B. J. Direct Submission of BAC-end sequences from Anopheles gambiae unpublished (2001)
JOURNAL	Other_GSSS: AG-ND-14112.TR
COMMENT	

1 (bases 1 to 721)
Shetty J., Malek J., Koo, H., Collins, F., Gardner, M. and Loftus, B.J.
Direct Submission of BAC-end sequences from *Anopheles gambiae*
Unpublished (2001)
Other_GSSS: AG-ND-14112.TR
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjoftus@tigr.org
This clone is from an *A. gambiae* BAC library (ND-TAM) provided by
F.H. Collins and sequenced by The Institute for Genomic Research
(TIGR). The BAC library was generated from *A. gambiae* PST strain
DNA. All DNA was extracted from newly hatched first instar larvae
to minimize the inclusion of DNA from microorganisms that inhabit
the gut. The DNA is derived from mixed sexes of larvae. The BAC
library was constructed at Texas A&M University BAC Center
University College Station, Texas 77843-2123, USA using a HindIII
partial digest.
seq primer: M13for

FEATURES	Class: BAC ends.	Location/Qualifiers
SOURCE	1..721	
	/organism="Anopheles gambiae"	
	/strain="PEST"	
	/db_xref="taxon:7165"	
	/clone="AG-ND-14112"	
	/clone_11b="ND-TAM"	
	/note="Vector: pCIBAC1; Site_1: HindIII"	
BASE COUNT	269 a 122 c 147 g 183 t	
ORIGIN		
Query Match	1.9% Score 41.2; DB 12; Length 721;	
Best Local Similarity	47.3% Pred. No. 2.6; Mismatches 138; Indels 0; Gaps 0;	
Matches 124; Conservative 0;		
QY 924	GAAGAACCCCTGACGCGAGATCAAGCTACCCGCTATCAATATGCGACAGATATGCTCA 983	
Db 28	GAGAAAGCATATGATTAAGAGATGCTACCAAGATTAGCTTAGCGGTCTAACTATGCAAG 87	
QY 984	GAGGTCTACTATTTGGAAGAATTCATGCTGATGTAACCGCGCTTCGCTTGTGACGT 1043	
Db 88	TGTTGCTACTATTTGGAAGAAACAGACCGCGGAACATTTGAAGCGGGCTTCAAAAACGTATC 147	
QY 1044	GATGAGCTGTAAAGCGTGGCGGAGAAAGACATTCGCGAGCTGATCCGTAACAGAGGCA 1103	
Db 148	AATTGCTACAAAAAAGAGTTGAGAAAAAATACACAAATGGCGCTGATAAAGCTGAAAA 207	
QY 1104	GAGTGTCTCTATGCGCATGATTTCTTCTCGMAAAGCGCTTATTAAGAGAGAGCCAA 1163	
Db 208	CAAGCTCTTTATGTAATATGTAATGCAAAATACATGATGATGCTACATCAAGCAATTTCTTA 267	
QY 1164	GCCACACCGTGATGACTTAT 1185	
Db 268	CAGAAATATCGAGAAAAATTAT 289	
RESULT 12		
LOCUS	BH371846 822 bp DNA linear GSS 10-DEC-2001	
DEFINITION	AG-ND-162M17.TF ND-TAM Anopheles gambiae genomic clone AG-ND-162M17	
ACCESSION	BH371846	
VERSION	BH371846.1 GI:17317971	
KEYWORDS	GSS.	
SOURCE	African malaria mosquito.	
ORGANISM	Anopheles gambiae	
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;	
	Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoides	
	; Anophelinae.	
REFERENCE	1 (bases 1 to 822)	
AUTHORS	Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J	
TITLE	Direct Submission of BAC-end sequences from Anopheles gambiae	
JOURNAL	Unpublished (2001)	
COMMENT	Other_GSSs: AG-ND-162M17.TR	
	Contact: Brendan J Loftus	
	Department of Eukaryotic Genomics	
	The Institute for Genomic Research	
	9712 Medical Center Dr., Rockville, MD 20850, USA	
	Tel: 301 838 0208	
	Fax: 301 838 3543	
	Email: bjloftus@tigr.org	
	This clone is from an A. gambiae BAC library (ND-TAM) provided by	
	F.H. Collins and sequenced by The Institute for Genomic Research	
	(TIGR). The BAC library was generated from A. gambiae PEST strain	
	DNA. All DNA was extracted from newly hatched first instar larvae	
	to minimize the inclusion of DNA from microorganisms that inhabit	
	the gut. The DNA is derived from mixed sexes of larvae. The BAC	
	library was constructed at Texas A&M University BAC Center	
	University, College Station, Texas 77843-2123, USA using a HindIII	
	partial digest.	
	Seq primer: M13 For	
	Class:BAC ends.	

FEATURES	Location/Qualifiers
SOURCE	1. 822
	/organism="Anopheles gambiae"
	/strain="PEST"
	/db_xref="taxon:7165"
	/clone="AG-ND-162M17"
	/clone_lib="ND-TAM"
	/note="Vector: pECBAC1; Site_1: HindIII"
BASE COUNT	235 a 142 c 155 g 290 t
ORIGIN	
Query Match	1.8%; Score 39.4; DB 12; Length 822;
Best Local Similarity	65.2%; Pred. No. 9;
Matches	58; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
QY 2007	AGCTATGAGTGGTGCACATCGAATTCGACCGCATCTCGAGCGCACAAATCAGCGTGGACAT 2066
Db 1	AGCTTTAAGCGGTGACATGTTTGGACCTAAATATACAAAGAACGATTAACTAGACGT 60
QY 2067	CCGCTACGTTCTCTTCATGATTCGACAAT 2095
Db 61	TAGATACGCTACTTGGGTAAATCGACAAGT 89
RESULT 13	
LOCUS	BH448937 687 bp DNA linear GSS 12-DEC-2001
DEFINITION	BOGZL677R BOGZ Brassica oleracea genomic clone BOGZL67, DNA sequence.
ACCESSION	BH448937
VERSION	BH448937.1 GI:17634648
KEYWORDS	GSS.
SOURCE	Brassica oleracea.
ORGANISM	Brassica oleracea
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica. 1 (bases 1 to 687)
AUTHORS	Town,C.D., Van Aken,S., Uterback,T. and Fraser,C.M.
TITLE	Whole genome shotgun sequencing of Brassica oleracea
JOURNAL	Unpublished (2001)
COMMENT	Other GSSs: BOGZL67TF Contact: Chris Town TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA. Tel: 301-838-3523 Fax: 301-838-0208 Email: cdtown@tigr.org DNA is from a doubled haploid provided by Tom Osborn. Seq primer: TR Class: sheared ends. Location/Qualifiers 1. 687 /organism="Brassica oleracea" /strain="T01000DH3" /db_xref="taxon:3712" /clone="BOGZL67" /clone_lib="BOGZ" /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT	209 a 117 c 159 g 202 t
ORIGIN	
Query Match	1.8%; Score 38.6; DB 12; Length 687;
Best Local Similarity	53.7%; Pred. No. 14;
Matches	80; Conservative 0; Mismatches 69; Indels 0; Gaps 0;
QY 1457	TGCGACAAGAGTGGTTCCTATAGCGACAAGTTCATCGCATGCTCAAGTCCATGAGACA 1516
Db 685	TCGACACTTTTGGGATCAAAAGACGACAGTTCAGATCCAAAGATCTAGTTAAGC 626
QY 1517	AGGAAAAGTTTCCCAAGGCTATCGAGAAAGATCCGGCAGTAGAAGCTTTCCAAAGACGTPAA 1576

Db 625 AGTGAACATTCACATCGCTATCATGATTAAGAGATTGAGCTTAATATATCATGC 566

OY 1577 TAGCTGCTGCTCGGCTATTCAGGCCGAT 1605

Db 565 ATGCAACTGTAGGATCTATAGGATTAT 537

RESULT 14

LOCUS AM375157 184 bp mRNA linear EST 04-FEB-2000

DEFINITION OY2-CT0121-280999-001-e01 CT0121 Homo sapiens cDNA, mRNA sequence.

ACCESSION AM375157

VERSION AM375157.1 GI:6879811

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 184)

AUTHORS HCGP <http://www.ludwig.org.br/ORESTES>.

TITLE The FAPESP/LICR Human Cancer Genome Project

JOURNAL Unpublished (1999)

COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(<http://www.ludwig.org.br/scripts/gethtml2.pl?tl-QV2t2-QV2-CT0121-280999-001-e01&t3-1999-09-28&t4-1>)
Seq primer: puc 18 forward
High quality sequence start: 18
High quality sequence stop: 88.
Location/Qualifiers
1. 184
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CT0121"
/dev_stage="Adult"
/note="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT 29 a 53 c 58 g 44 t

ORIGIN

Query Match 1.88; Score 38.4; DB 9; Length 184;
Best Local Similarity 51.18; Pred. No. 7.9;
Matches 90; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

OY 612 CAGCTCTGTAGTAACTTCGAGCGGATACGACAACTGATGTGCGCGCTACACAGG 671

Db 8 CAGCTTTCTCTCAGCTGCTTGGGCTCTGGAATTTGGGGGGGCGCTCTCTGAT 67

OY 672 CGACTTCAGCGTATTCGCGGTATGCGGTCCGACACCGCGCGCGAATACAGCA 731

Db 68 CAACCTTCCACGAGAGCTCTGTGCGTTGGGGAACAGCACCTCTGTCAGGCA 127

OY 732 GGACATAAACCCTTAAGCCGTTACTTGGCTGCCATATCATGCAAGGCTACA 787

Db 128 GGCCAGATGCCCGGGGGGCGTCTCTCATGAGTGCATGAGGCGTGCA 183

RESULT 15

BF426788 466 bp mRNA linear EST 18-APR-2001

LOCUS BF426788

DEFINITION dF72a12.v1 Xenopus laevis unfertilized egg cDNA library Xenopus laevis cDNA clone IMAGE:3744694 5' similar to TR:093308 093308 14S COHESIN SMC1 SUBUNIT. ;, mRNA sequence.

ACCESSION BF426788

VERSION BF426788.1 GI:11438250

KEYWORDS EST.

SOURCE African clawed frog.

ORGANISM Xenopus laevis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae; Xenopodinae; Xenopus.

REFERENCE 1 (bases 1 to 466)

AUTHORS Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Page,D., Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person ,B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R., Waterston,R. and Wilson,R.

TITLE Washu Xenopus EST project, 1999

JOURNAL Unpublished (1999)

COMMENT Contact: Sandy Clifton, Ph.D.
Washu Xenopus EST project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
Library constructed by Bruce Blumberg
Library normalized by Jihwan Song
DNA sequencing by: Washington University Genome Sequencing Center
Clone distribution: Xenopus clone distribution information for this library can be found through Research Genetics, visit their web page at: <http://www.resgen.com/> Please reference the id listed below when ordering this clone: Source lab clone id - xlineg006b23
Seq primer: -40RP from Glbco
High quality sequence stop: 407.
Location/Qualifiers
1. 466
/organism="Xenopus laevis"
/db_xref="taxon:8335"
/clone="IMAGE:3744694"
/clone_lib="Xenopus laevis unfertilized egg cDNA library"
/tissue_type="unfertilized egg"
/lab_host="Top-10 F."
/note="Vector: plasmidscript SK-; Site_1: EcoRI; Site_2: XhoI; cDNA was prepared from 2ug of poly A+ RNA. EcoRI-XhoI cut cDNA was then ligated into Unizap-XR (Stratagene) with EcoRI at the 5' end and XhoI at the 3' end. SS-library phagemids were prepared by mass excision from the original library and normalized by hybridization to biotinylated driver (prepared from the same library by PCR) to Cot-omega of 11. After removal of hybrids and excess driver by streptavidin sepharose chromatography, the ss-phagemids were made double stranded and electroporated into Top-10 F'. Original library construction by Bruce Blumberg (Blumberg et al., 1991 Science 253, 194-196; Hawist et al., 1995, Genes Dev. 9, 2923-2935). Note: This is a Xenopus Gene Collection (XGC) library."
BASE COUNT 184 a 80 c 131 g 71 t

ORIGIN

Query Match 1.88; Score 38.4; DB 10; Length 466;
Best Local Similarity 56.28; Pred. No. 13;
Matches 72; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

OY 391 AAGATCGTAGGATTAAGCAGGTAAAGACGCTCAAGGGATGATCATGACGATG 450

Db 105 AAGGACAGAAGCATATGACAAAGTGAGGAGAGCTCAAGGACAGACAAAGATTG 164

OY 451 GAGCGTGCAGCAAGCTCAGAGGTATGCCAAGAACTGCGCAAAAAGAAATGACAG 510

Db 165 GGAAGATGATGACAGACAGCGCATTGAGAAAGATCAAAAGAGAGATGCACAA 224

OY 511 GAGAACCA 518

Db 225 CTGAATCA 232

Search completed: October 17, 2002, 21:37:38
Job time : 1743 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 17, 2002, 21:35:21 ; Search time 60 Seconds
(without alignments)
1318.076 Million cell updates/sec

Title: US-10-008-355-2

Perfect score: 3719
Sequence: 1 MDMKLSILGALLGASG.....LFMDKMGCCPLDLEKLI 712

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues
Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	133	3.6	883	22	AAU37789
2	130	3.5	883	20	AAU08339
3	130	3.5	883	21	AAU90514
4	127.5	3.4	1194	18	AAW21725
5	127	3.4	883	22	AAU38091
6	123.5	3.3	713	20	AAU43380
7	122.5	3.3	800	18	AAW21723
8	122.5	3.3	813	18	AAW21728
9	122.5	3.3	1181	18	AAW21727
10	122.5	3.3	1194	18	AAW21724
11	122.5	3.3	1194	18	AAW21726

12	117.5	3.2	951	20	AAV34536
13	117.5	3.2	953	20	AAV34403
14	114	3.1	711	17	AAU88649
15	113	3.0	708	11	AAU43567
16	113	3.0	2042	19	AAW56319
17	112.5	3.0	907	22	AAU78650
18	112	3.0	724	22	ABG15384
19	112	3.0	724	22	ABG17531
20	110.5	3.0	1027	14	AAU42203
21	110.5	3.0	1027	14	AAU43699
22	110	3.0	705	17	AAU88645
23	109.5	2.9	845	21	AAU90942
24	109	2.9	416	21	AAU34432
25	109	2.9	416	21	AAU34433
26	109	2.9	1577	17	AAU91047
27	109	2.9	2234	21	AAU81502
28	108.5	2.9	950	21	AAU42742
29	108.5	2.9	950	22	AAU93087
30	108	2.9	2027	22	ABG07898
31	108	2.9	4536	20	AAU96826
32	107.5	2.9	655	20	AAU49226
33	107.5	2.9	655	20	AAU32099
34	107.5	2.9	694	21	AAU81653
35	106.5	2.9	523	22	AAU36325
36	106.5	2.9	1162	22	ABG09158
37	106.5	2.9	1644	22	ABG12176
38	106	2.9	918	22	ABU59819
39	106	2.9	972	14	AAU51281
40	106	2.9	1592	14	AAU32925
41	106	2.9	2721	8	AAU70647
42	105.5	2.8	403	17	AAU40855
43	105.5	2.8	609	14	AAU88657
44	105.5	2.8	710	22	ABU62130
45	105	2.8	361	22	AAU60881

ALIGNMENTS

RESULT 1	AAU37789	standard; Protein; 883 AA.
ID	AAU37789	
AC	AAU37789	
XX		
DT	14-FEB-2002	(first entry)
XX		
DE	Streptococcus pneumoniae cellular proliferation protein #218.	
XX		
KW	Antisense; prokaryotic cellular proliferation protein; antibiotic; antibacterial; drug design.	
XX		
OS	Streptococcus pneumoniae.	
XX		
PN	W0200170955-A2.	
XX		
PD	27-SEP-2001.	
XX		
PF	21-MAR-2001; 2001WO-0509180.	
XX		
PR	21-MAR-2000; 2000US-191078P.	
PR	23-MAY-2000; 2000US-206848P.	
PR	26-MAY-2000; 2000US-207727P.	
PR	23-OCT-2000; 2000US-242578P.	
PR	27-NOV-2000; 2000US-253625P.	
PR	22-DEC-2000; 2000US-257931P.	
PR	16-FEB-2001; 2001US-269308P.	
XX		
PA	(ELIT-) ELITRA PHARM INC.	
XX		
PI	Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ,	
PI	Yamamoto RT, Xu HH;	
XX		

Phosphorylase kinase
Phosphorylase kinase
Nucleoside triphosphate
Human cancer assoc
Haemophilus paraga
XRN-100, unidentified
Novel human diago
Novel human diago
Protein L. Peptoc
Protein L. Peptoc
Nucleoside triphosphate
Cenarchaeum symbio
Gene 46 human ser
Human secreted pro
Alpha-D-glucosyltr
Streptococcus pneu
Human ORF ORF2506
Human protein sequ
Novel human diago
Amino acid sequenc
CbpA of serotype 4
Choline binding pr
Streptococcus pneu
Pseudomonas aerugi
Novel human diago
Novel human diago
Drosophila melanog
Helminth aminopept
Glucosyltransferase
Sequence of N-term
38kd regression as
N.meningitidis B21
Drosophila melanog
Mevalonate pathway

DR WPI: 2001-611495/70.
DR N-PSDB: AAS55648.

PR New polynucleotides for the identification and development of
XX antibiotics, comprise sequences of antisense nucleic acids -
XX
XX
PS Example 3; Seq ID No 13382; 511pp; English.

XX The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhimurium, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
XX

Sequence 883 AA:

Query Match 3.6%; Score 133; DB 22; Length 883;
Best Local Similarity 20.4%; Pred. No. 0.025;
Matches 104; Conservative 65; Mismatches 166; Indels 176; Gaps 24;

QY 192 DVF-----KDVRYVFPSSVSGKFGGDTNNMMPRTTGFVSFRVYAGADNRPAREYS 243
DB 26 DVFKPSGQOKAPYSIVIPPNVTGKL-----HLG-----HA 57
QY 244 KDNKPYKPYFAAVSMOGKADYAMTIGFPGSTDRYL-TSGVGEDRIENNNPREVNG 302
DB 58 WDT-TLQDIIIRQKRMQFDP-----TLMLPGMDHAGIATQAKVEERLNGEGISRDLGR 110
QY 303 ----IKQIGWKEMASDAQTRIKYA-SKYAQSANTYKNSIGNNRGLARLDVIGRRKRAER 357
DB 111 ESFLTKRWEMKD-----EYATYTIKQMGKGLSVDSRERFTLDEGLS-----KAVRK 158
QY 358 APADMIRK-----NGKSAYVGD-----VLSLEKAYK-----EGAKANREMT 394
DB 159 VFDYLYKKWIRGEFTIIMDPAARTALSDIEVIRKQVGAFTYHNNYMLEDSRLAEVAT 218
QY 395 VLSLEFGTEVVRFAQFANALATNPDAHAGILKSLDKYKDYLPs-----LDRKVLPM 449
DB 219 TRPETMFGDV-----ALAVNPE-----DPRYKDLIGKNVILPLANKLPIV 259
QY 450 LD-----IVRRIRPADKLPDIFKKNVIDKKFKGDTKKYADVFEDKSVVPYSDK 496
DB 260 GDEHADPERGTGVKITTPAHDPNDFLVQGRHNLPOVNNIMNDGTMMNDALEFESGM---DR 316
QY 497 FHAMLKSMQKEFAKAIEKDPAVELSKSYIAARA-----IQAD 535
DB 317 FEAR-----KAVYAKLEELGALVKIEKRVHSVGHSERTGVVVEPRLSITQWFKMQDLAKN 371
QY 536 AMANAYAIKGRFLFFAGIREMYPGRALPS-----DANFTYKMSYGSITKYEPODG 586
DB 372 ALANDQTECK-----VEFYPRFNDFTLQWMMENVDWYISQGLMMGH-----QIP 416
QY 587 AMYNTHTTGGKVLKODPKSDEPAVOENITLD 617
DB 417 AMYN--ADGEMTVYGEARPEGDGTODEVDLD 445

RESULT 2
AAV08339

ID AAV08339 standard; Protein; 883 AA.

AC AAV08339;

XX 19-JUL-1999 (first entry)

DE S. pneumoniae vals protein.

KW vals; treatment; antagonist.

OS Streptococcus pneumoniae.

PM JP11113577-A.

PD 27-APR-1999.

PF 17-OCT-1997; 97JP-0321886.

PR 17-OCT-1997; 97JP-0321886.

PA (SMIK) SMITHKLINE BEECHAM CORP.

DR WPI: 1999-320830/27.

DR N-PSDB: AAX57027.

PR New vals polypeptide and polynucleotide - useful for treatment of
XX conditions associated with abnormal vals protein levels
XX
XX

PS Claim 1a; Page 25-27; 27pp; Japanese.

CC This invention describes the isolation of a novel Streptococcus
CC pneumoniae vals protein. The new polypeptide and antagonist are
CC useful for the treatment of an individual with abnormal vals
CC protein levels, by administering vals to individuals requiring
CC vals polypeptide, and administering the antagonist to individuals
CC requiring inhibition of vals polypeptide.
XX
XX

Sequence 883 AA:

Query Match 3.5%; Score 130; DB 20; Length 883;
Best Local Similarity 20.4%; Pred. No. 0.046;
Matches 104; Conservative 63; Mismatches 166; Indels 176; Gaps 24;

QY 192 DVF-----KDVRYVFPSSVSGKFGGDTNNMMPRTTGFVSFRVYAGADNRPAREYS 243
DB 26 DVFKPSGQOKAPYSIVIPPNVTGKL-----HLG-----HA 57
QY 244 KDNKPYKPYFAAVSMOGKADYAMTIGFPGSTDRYL-TSGVGEDRIENNNPREVNG 302
DB 58 WDT-TLQDIIIRQKRMQFDP-----TLMLPGMDHAGIATQAKVEERLNGEGISRDLGR 110
QY 303 ----IKQIGWKEMASDAQTRIKYA-SKYAQSANTYKNSIGNNRGLARLDVIGRRKRAER 357
DB 111 ESFLTKRWEMKD-----EYATYTIKQMGKGLSVDSRERFTLDEGLS-----KAVRK 158
QY 358 APADMIRK-----NGKSAYVGD-----VLSLEKAYK-----EGAKANREMT 396
DB 159 VFDYLYKKWIRGEFTIIMDPAARTALSDIEVIRKQVGAFTYHNNYMLEDSRLAEVAT 218
QY 397 --SEFLFGTEVVRFAQFANALATNPDAHAGILKSLDKYKDYLPs-----LDRKVLPM 449
DB 219 TRPETMFGDV-----ALAVNPE-----DPRYKDLIGKNVILPLANKLPIV 259
QY 450 LD-----IVRRIRPADKLPDIFKKNVIDKKFKGDTKKYADVFEDKSVVPYSDK 496
DB 260 GDEHADPERGTGVKITTPAHDPNDFLVQGRHNLPOVNNIMNDGTMMNDALEFESGM---DR 316
QY 497 FHAMLKSMQKEFAKAIEKDPAVELSKSYIAARA-----IQAD 535
DB 317 FEAR-----KAVYAKLEELGALVKIEKRVHSVGHSERTGVVVEPRLSITQWFKMQDLAKN 371
QY 536 AMANAYAIKGRFLFFAGIREMYPGRALPS-----DANFTYKMSYGSITKYEPODG 586

DB 372 AIANODEDK-----VEFYPPRFNDTFLQWMEVHDWVISRQLMNGH-----QIP 416

QY 587 AMYNHTTGKGVLEKODPKSDEFAVOENIID 617

DB 417 AMYN--ADGEMVGEBAPEGDGWTODEPYLD 445

RESULT 3

AA90514

ID AA90514 standard; Protein; 883 AA.

AC AA90514;

XX

DT 15-AUG-2000 (first entry)

DE Streptococcus pneumoniae valyl tRNA synthetase (vals).

XX

KW Valyl tRNA synthetase; vals; inhibitor; stringent response;

KM drug screening; antibacterial; antibiotic; genetic immunisation;

KW antibody; bacterial infection; meningitis.

XX

OS Streptococcus pneumoniae strain 0100993.

XX

XX US6051413-A.

PN

PD 18-APR-2000.

XX

PF 24-SEP-1998; 98US-0159539.

XX

PR 18-APR-1996; 96GB-0007791.

PR 17-OCT-1997; 97US-0953492.

PR 18-APR-1997; 97US-0844064.

XX

PA (SMK) SMITHKLINE BEECHAM CORP.

XX

PI Brown JR, Lawlor EJ, Wang M, Jaworski DJ;

XX

DR WPI: 2000-338311/29.

DR N-PSDB; AAA14369.

XX

PT Novel vals polypeptides of valyl tRNA synthetase family useful for

PT treating otitis media, conjunctivitis, pneumonia and bacteremia

PT comprises a specified amino acid sequence -

XX

PS Claim 5; Columns 9-10; 21pp; English.

XX

CC This sequence represents Streptococcus pneumoniae valyl tRNA synthetase

CC (vals, NCIMB Deposit No. 40794). Streptococci are the cause of several

CC types of human diseases, including otitis media, conjunctivitis,

CC pneumonia, bacteraemia, sinusitis, pleural empyema, endocarditis and

CC especially meningitis. The frequency of S. pneumoniae infections has

CC risen dramatically over the past 20 years, probably due to the emergence

CC of multiply antibiotic resistant strains and an increasing population of

CC immunocompromised people. Vals represents a target for new antibacterial

CC agents. Inhibition of tRNA synthetases such as vals leads to a reduction

CC in the levels of charged tRNA, which triggers a cascade of responses

CC (known as the stringent response) resulting in a state of dormancy in the

CC bacterium. Vals, its variants and fragments, anti-vals antibodies, vals

CC inhibitors and nucleotides encoding vals may be used in the diagnosis,

CC prevention and treatment of bacterial infections such as meningitis.

CC Vals can be used to screen compounds for inhibitory activity. Vals

CC may also be useful as an antigen for vaccination of a host to produce

CC specific antibodies which protect against bacterial invasion into

CC damaged tissues. Such antibodies could, for example, prevent the

CC adherence of bacteria to wounds. Nucleotides encoding vals may be used as

CC diagnostic reagents and therapeutic or prophylactic agents, particularly

CC for genetic immunisation.

XX

SO Sequence 883 AA;

Query Match 3.5%; Score 130; DB 21; Length 883;

Best Local Similarity 20.4%; Pred. No. 0.046;

Matches 104; Conservative 63; Mismatches 168; Indels 176; Gaps 24;

QY 192 DVF-----KQVRYFAPSPSSVGKFGGDTDMMMPRHTGDSVRYVYAGADNRPAREYS 243

DB 26 DVFPGSGDOKAKPYSIVIPPNVTKL-----HLG-----HA 57

QY 244 KDNKPYKPYFAVNSMGYKADYAMTIGPGSTDRYL-TSMGVEDRIENENPRIEVRG 302

DB 58 WDT-TLQDIIIRKRMQGF-----TWLPGMDHAGIATQAKYERLRKRGISRIIDLGR 110

QY 303 ----IKQIWKEMASADOATRIRYA-SKYAQSANYKNSIGMNRGLARLDVIGRKRAEER 357

DB 111 ESFLTKVMEWKD---EYATTIKEQMGKMLSDVYSRERTPLEGLS-----KAVRK 158

QY 358 AFDADIRK-----NGSAVYGD-----VLSLEAYKEGAKANEMRYL----- 396

DB 159 VFVDLYRKGMVYRGEFTIMDPAARTALSDIEVIHKDVEGAFYHNMVMEJEDGSRLVEAT 218

QY 397 --SETLFGGTEVVRFAOFANALATNPDAHGIKLSDDKYKDYLPs-----LDKRYLPAM 449

DB 219 TRETFMGDV-----AVAVNPE-----DPRYKDLIGNVILPIANKLIPY 259

QY 450 LD-----IVRRIRPADKLPIFKNVIDKKFKGDTKKYVDYFVDSKVPSYSDK 496

DB 260 GDEHADPEFGTGVYKITPAHPDNPDLVQGRHNPQYVVMNDGTMTNDLAFEEFGM---DR 316

QY 497 FHAMLSMDKEKFAKALENDPAVELSKSYIAARA-----IQAD 535

DB 317 FEAR-----KAVAKLEEALVKIEKRVAHSVGSERTGVYVERPRLSTQFVKMDLAKN 371

QY 536 AMANAYAIKGRKLPFAGLREMYGRALPS-----DANPTMNSYGSIGKYPQDG 586

DB 372 AIANODEDK-----VEFYPPRFNDTFLQWMEVHDWVISRQLMNGH-----QIP 416

QY 587 AMYNHTTGKGVLEKODPKSDEFAVOENIID 617

DB 417 AMYN--ADGEMVGEBAPEGDGWTODEPYLD 445

RESULT 4

AAW21725

ID AAW21725 standard; protein; 1194 AA.

XX

AC AAW21725;

XX

DT 01-OCT-1997 (first entry)

XX

DE Modified streptokinase, rSK6mut.

XX

KW Plasminogen-binding fragment; streptokinase; degradation; MBP;

KW thrombolytic agent; blood clot; bolus; maltose-binding protein.

XX

OS Streptococcus equisimilis.

XX

XX

FH Key Location/Qualifiers

FT Protein 1..381

FT Protein /label= Maltose-binding_protein

FT Protein /note="acts as blocking group"

FT Protein 382..1194

FT Protein /label= Modified_streptokinase

FT Misc-difference 391

FT Misc-difference 417 /label= R10A

FT Misc-difference 417 /label= R36A

FT Misc-difference 426 /label= R45A

FT Misc-difference 432 /label= R51A

FT Misc-difference 440 /label= R59A

FT Misc-difference 460 /label= R59A

FT Misc-difference 766 /label= K385A

FT /note="Given in the specification as K386A"

XX W09641883-A1.
XX 27-DEC-1996.
XX 07-JUN-1996; 96WO-US09640.
XX 09-JUN-1995; 95US-0488940.
XX (HARD) HARVARD COLLEGE.
XX
XX Read GL;
XX WPI; 1997-065469/06.
XX Modified forms of streptokinase resistant to enzymatic cleavage -
PT useful as thrombolytic agents in treating thrombosis and in medical
PT equipment
XX
XX Claim 16; Page 44-48; 65pp; English.
XX
XX This sequence represents a fusion protein between maltose-binding
CC protein and a modified form of the plasminogen-binding fragment of
CC streptokinase containing 6 point mutations. This modified streptokinase
CC has an in vitro degradation rate at least 2 times slower than that of
CC native streptokinase. Compounds containing modified streptokinases
CC are specifically used as thrombolytic agents for dissolving blood
CC clots in vivo in a mammal, preferably at a dose of 20000 U/kg,
CC optionally as a bolus rather than by continuous infusion.
XX
XX Sequence 1194 AA;
XX

Query Match 3.4%; Score 127.5; DB 18; Length 1194;
Best Local Similarity 20.1%; Pred. No. 0.12; Mismatches 263; Indels 167; Gaps 35;

Matches 133; Conservative 99; Mismatches 263; Indels 167; Gaps 35;

QY 75 GIVSDQGL-----IFTNHCYGAIQSQSTVDHDLRDGFVSRMGELPIP--GLS 125
DB 571 GDTISQELLAQAOSILKNHPGYTIERDSST--VTHNDIFRTI--LPMDEFTYR 624
QY 126 VKYLAKIYKVDKVEGOLKGTDEMERLKAQVCOELAKKE----- 167
DB 635 VKNRBOAIRINK-----SGLINEINNTDLISEKYVYLKGGKPYDPDRSHLKLFTIKY 679
QY 168 -NADENQCIYEPFYSNNY--FLIVYDFEKDVRMFAFPSSVCKFG-----GDTDN 216
DB 680 VDVQNELLSQQLTASERNLDFBDLYDRPKAKLY--NNLDAFGIMDTLTGXVED 736
QY 217 WKMPRHGDSFVFRVYAGADNRPAEYSKDNKPKPVYFAVSMOGYKADYAMTIGPGS 276
DB 737 ---NHDDTNRIITYVMG--KRP--EGENASYH-----LAYDADRYEEREYVS 778
QY 277 TDRYLTSNGVEDRINENNPR-----EVNGIKGIVKEMASDAQARIKASKYAC-- 328
DB 779 YLRY-TGPIPIPNPDKNNSQLVSVAGTEGTNDISLKEFEIDLTSRPAIGGTEGGL 837
QY 339 ---SANYKNSIGMNGRLARLDVIGRRAEERAFADWIRKN-----GKSAYYGV 376
DB 838 SPSKAPFATDSGAMGHKLEKADLL--KAIQEQLIAN-VHSNDYFEVIDFASDATITBRN 894
QY 377 SLEKAYVEGAKANREMYTLSTLFEGTEVVRFAOPANALATNPANAGILKSLDDKY-- 434
DB 895 GKVYFADKDG--VLTPTPOVQEFLLSGHVRVRYKE-----KPIYONQ--KSVDEYTV 944
QY 435 -----KDYLPPL--DRKVLPLAMLDIVRRIPADKLPDIFKAVIDKKFGGTTRYADPV 485
DB 945 QPTPLNPDDDFRPGKLKLTLL--AIGDITTSQELLAQAOSILKNHPG---YTYIE 998
QY 486 FDKSVVPYSDEKHAMLKSDKEKFAKAEKDPAVELSKVAARAADAMANA--YAE 544
DB 999 RRSSTIVTHNDIFRTILPMDQFTYRVKNRQAYRIKKSGLNEFINMTDLISEKYVYLK 1058
QY 545 KGK-----RLFFRAGLREMTPGRALPSDANFT---MMMSGSTIKGYEPDQGA--W 588

DB 1059 KGEKPYDPDRSHLKLFTIKYVDVDTNELLSKSEQLLTASERNLDFRDL--YDPRKAKLL 1116
QY 589 YN-----YHTGKGKVLKQDPKSDFAVQENIILFR--FKNYGR--YAEQGQJHI 635
DB 1117 YNNLDAFGIMDTYLTGK-----VEDNHDDTNRIITYVMGKRPEGENASYHL 1162
QY 636 AF 637
DB 1163 AY 1164
RESULT 5
ID AU38091
XX AU38091 standard; Protein; 883 AA.
AC AU38091;
XX
XX 14-FEB-2002 (first entry)
DE Streptococcus pneumoniae cellular proliferation protein #520.
XX
XX Antisense: prokaryotic cellular proliferation protein;
KW antibiotic; antibacterial; drug design.
XX OS Streptococcus pneumoniae.
XX PN W0200170955-A2.
XX PD 27-SEP-2001.
XX PF 21-MAR-2001; 2001WO-US09180.
XX PR 21-MAR-2000; 2000US-191078P.
XX PR 23-MAY-2000; 2000US-206848P.
XX PR 26-MAY-2000; 2000US-207727P.
XX PR 23-OCT-2000; 2000US-242578P.
XX PR 27-NOV-2000; 2000US-253625P.
XX PR 22-DEC-2000; 2000US-257931P.
XX PR 16-FEB-2001; 2001US-269308P.
XX PA (ELIT-) ELITRA PHARM INC.
XX PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX WPI: 2001-611495/70.
XX DR N-PSDB; AAS55950.
XX PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
XX Example 3; Seq ID NO 13684; 51pp; English.
XX
XX The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Streptococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.

Db	432	RAHGGKTEOGISPKSPKPRATDSGAMSHLEKADLL--KAIOEOLIAN--VHSDVFEYI	488
Qy	319	RIKYASKAAQ-----SANYKWSIGNNRGLARLDVIGRRRAEERAFADMIKRN-----	366
Db	372	KORIYEEEREYVSYLRYITCTPIPDNPNDKNNQIYVSVAGYEGTNGQDISLFFELDLIS	431
Qy	277	TDRY-----LTSWGEYDRIENENNRI-----EYRGIKOGIWKESADQAT	318
Db	343	----NHDDTNRITTYIMG--KRP-----EGENASYHLAY-----D	371
Qy	217	WMPPRHGTGDFSEVRVYAGADNRPAPKSKNPKPYKPYFAAVSMQGYKADYAWTIGFPCS	276
Db	286	VDDVDELILKSSQLTLASERINLDFRDLVDPDKAKLLY--NNLDAFGIMDYTLTGKVED	342
Qy	168	NADENOLCIVEPFYSNNXY---FLIYDVYFADVNRVFAFSPSYGKFG-----GDTDN	216
Db	231	YKNRQAVRINKK-----SGLNDEINMTDLISEKYYVLKKGKGPYDPDRSHLKLFTIKY	285
Qy	126	VKYLKRYKVDYDKVBSQLKGTDEMERLKAQVCOELAKK-----	167
Db	177	GDITRSQELLAQAQSLINKNHPGYTIYERDSSI---VTHDNDIIFRI---LPMDEFTYR	230
Qy	75	GTVSDQGL-----IFTNHCYGGAISQSTVDDHDLRDFGVFSRTMGELPIP--GLS	125
Db	130	Matches 130; Conservative 99; Mismatches 254; Indels 191; Gaps 35	
Qy	3	Query Match 3.3%; Score 122.5; DB 18; Length 800;	
Db	130	Best Local Similarity 19.3%; Pred. No. 0.18;	
Qy	3	Sequence 800 AA;	
Db	130	Matches 130; Conservative 99; Mismatches 254; Indels 191; Gaps 35	
Qy	3	Query Match 3.3%; Score 122.5; DB 18; Length 800;	
Db	130	Best Local Similarity 19.3%; Pred. No. 0.18;	
Qy	3	Sequence 800 AA;	
Db	130	Matches 130; Conservative 99; Mismatches 254; Indels 191; Gaps 35	
Qy	3	Query Match 3.3%; Score 122.5; DB 18; Length 800;	
Db	130	Best Local Similarity 19.3%; Pred. No. 0.18;	
Qy	3	Sequence 800 AA;	
Db	130	Matches 130; Conservative 99; Mismatches 254; Indels 191; Gaps 35	
Qy	3	Query Match 3.3%; Score 122.5; DB 18; Length 800;	
Db	130	Best Local Similarity 19.3%; Pred. No. 0.18;	
Qy	3	Sequence 800 AA;	
Db	130	Matches 130; Conservative 99; Mismatches 254; Indels 191; Gaps 35	
Qy	3	Query Match 3.3%; Score 122.5; DB 18; Length 800;	
Db	130	Best Local Similarity 19.3%; Pred. No. 0.18;	
Qy	3	Sequence 800 AA;	
Db	130	Matches 130; Conservative 99; Mismatches 254; Indels 191; Gaps 35	
Qy	3	Query Match 3.3%; Score 122.5; DB 18; Length 800;	
Db	130	Best Local Similarity 19.3%; Pred. No. 0.18;	
Qy	3	Sequence 800 AA;	
Db	130	Matches 130; Conservative 99; Mismatches 254; Indels 191; Gaps 35	
Qy	3	Query Match 3.3%; Score 122.5; DB 18; Length 800;	
Db	130	Best Local Similarity 19.3%; Pred. No. 0.18;	
Qy	3	Sequence 800 AA;	
Db	130	Matches 130; Conservative 99; Mismatches 254; Indels 191; Gaps 35	
Qy	3	Query Match 3.3%; Score 122.5; DB 18; Length 800;	
Db	130	Best Local Similarity 19.3%; Pred. No. 0.18;	
Qy	3	Sequence 800 AA;	
Db	130	Matches 130; Conservative 99; Mismatches 254; Indels 191; Gaps 35	
Qy	3	Query Match 3.3%; Score 122.5; DB 18; Length 800;	
Db	130	Best Local Similarity 19.3%; Pred. No. 0.18;	
Qy	3	Sequence 800 AA;	
Db	130	Matches 130; Conservative 99; Mismatches 254; Indels 191; Gaps 35	
Qy	3	Query Match 3.3%; Score 122.5; DB 18; Length 800;	
Db	130	Best Local Similarity 19.3%; Pred. No. 0.18;	
Qy	3	Sequence 800 AA;	
Db	130	Matches 130; Conservative 99; Mismatches 254; Indels 191; Gaps 35	
Qy	3	Query Match 3.3%; Score 122.5; DB 18; Length 800;	
Db	130	Best Local Similarity 19.3%; Pred. No. 0.18;	
Qy	3	Sequence 800 AA;	
Db	130	Matches 130; Conservative 99; Mismatches 254; Indels 191; Gaps 35	
Qy	3	Query Match 3.3%; Score 122.5; DB 18; Length 800;	
Db	130	Best Local Similarity 19.3%; Pred. No. 0.18;	
Qy	3	Sequence 800 AA;	
Db	130	Matches 130; Conservative 99; Mismatches 254; Indels 191; Gaps 35	
Qy	3	Query Match 3.3%; Score 122.5; DB 18; Length 800;	
Db	130	Best Local Similarity 19.3%; Pred. No. 0.18;	
Qy	3	Sequence 800 AA;	
Db	130	Matches 130; Conservative 99; Mismatches 254; Indels 191; Gaps 35	
Qy	3	Query Match 3.3%; Score 122.5; DB 18; Length 800;	
Db	130	Best Local Similarity 19.3%; Pred. No. 0.18;	
Qy	3	Sequence 800 AA;	
Db	130	Matches 130; Conservative 99; Mismatches 254; Indels 191; Gaps 35	
Qy	3	Query Match 3.3%; Score 122.5; DB 18; Length 800;	
Db	130	Best Local Similarity 19.3%; Pred. No. 0.18;	
Qy	3	Sequence 800 AA;	
Db	130	Matches 130; Conservative 99; Mismatches 254; Indels 191; Gaps	

```

QY 367 -GKAVYGVDPVLSLEKRYKEGAKANRMETLSEFLPGGTVEVFAQAFANALATNPDAH 422
Db 489 DFASDATTIDRNGKRYFADKDG- VTLPTPOVQEFLLSGHVRHYKE-----KPIQ 540
QY 425 GILKSLDQRY-----KDYLP-DRKVLPAAMDIVRRRIPADKLPDIFKNVIDKK 472
Db 541 A--KSVDEYVYQFLPPLAPDDDFRGLKDTLLKTL--AIDDTITSGEELLAQOSILKN 596
QY 474 FKGDTRKYADVFVDKSVYPYSDKFAHMLKSMCKEKFAKALEIKDPAVELSKSVIAARA 533
Db 597 HPG-----VTIERDSSIVTVDNDIFRTLLPMDQETTYRVKNREDAVRIKKSGNLEE 652
QY 534 ADAMAANA-YALEIKR-----RLEFAGLEMYGRALPDPDANT---MMSYGS 576
Db 653 TDLISEKYYVLKKGKPYDPDRSHLKTFTLKVVDPVDNTELLKSEQLITASERNLDRDL 712
QY 579 KGYEPQDA--WYN-----YHTTGKGVLEKODPKSDEFAQOENITLDFR--TKNYG 625
Db 713 --YDRDKAKLLNNLDAFGIMDTLLTGK-----VEDNNDITNRIITYWVG 756
QY 626 R--VAENGOLHIAF 637
Db 757 KRPEGENASVHLAY 770

RESULT 8
AAW21728
ID AAW21728 standard; protein; 813 AA.
XX AC
XX AAW21728;
XX
XX 01-OCT-1997 (first entry)
XX
XX Wild type plasminogen-binding fragment of Streptokinase.
DE
XX
XX Plasminogen-binding fragment; streptokinase; degradation; MBP;
XX thrombolytic agent; blood clot; bolus; maltose-binding protein.
XX
XX Streptococcus equisimilis.
OS
XX
XX W09641883-A1.
XX PN
XX 27-DEC-1996.
XX PD
XX 07-JUN-1996; 96WO-US09640.
XX PF
XX 09-JUN-1995; 95US-0488940.
XX PR
XX
XX (HARD ) HARVARD COLLEGE.
XX PA
XX
XX Reed GL;
XX
XX WPI; 1997-065469/06.
XX DR
XX
XX Modified forms of streptokinase resistant to enzymatic cleavage -
XX useful as thrombolytic agents in treating thrombosis and in medical
XX equipment
XX
XX Example 1; Page 12-13; 65pp; English.
XX PS
XX
XX This sequence represents the wild type plasminogen-binding fragment
XX of streptokinase. This fragment was used in the design of a
XX modified streptokinase has an in vitro degradation rate at least
XX 2 times slower than that of native streptokinase. Compounds
XX containing modified streptokinases are specifically used as
XX thrombolytic agents for dissolving blood clots in vivo in a
XX mammal, preferably at a dose of 20000 U/kg; opt. as a bolus
XX rather than by continuous infusion.
XX
XX Sequence 813 AA;
SQ

```


Best Local Similarity 19.3%; Pred. No. 0.18;
Matches 130; Conservative 99; Mismatches 254; Indels 191; Gaps 35;

QY	75	GIATVSDGL-----IFTNHHCYGAIGOSQVYDHDYLRDGVSTRTMGEELPIP-----GLS	125
Db	190	GDITTSOELLAQAOSILNNKHHPGTYTYERDSSI---VTHDNDIFRTI---LPMODETYR	243
QY	126	VKYLRLKIVKTVYDVEGOLKGTIDEMERLKKAEVCOELAKR-----	167
Db	244	VKNNEQAYRINK-----SLNLEINNTDILSEKYYVLKGGKRPDPDPDRSHLKIPTYK	298
QY	168	-MADENOLIVEPFYSNNEY---FLIVDYEFKDYRMVEAFPPSSVGKFG-----GDTDN	216
Db	299	VDVDTNELTKSEQLLTASEENLDFRLDYPDRKAKILTY---NLDLAFGIMDYTLTGKVED	355
QY	217	WMPRHRTGDESVRYRYAGADNRAEFSKNNKYEKPYEFAAVSMOGYKADYAMTIGFPGS	276
Db	356	-----NHDDTNRITTYVYG--KRP--EGENASVHLAY-----D	384
QY	277	TDRY-----LTSWGVEDRIENENNPRI-----EYRGITKOGTWKAMSADAT	318
Db	385	KDRTEREEREVYSYLRTGTPRIDNPNNDKNNQOLVYSVAGIEGTINODISLKEFBIDLTS	444
QY	319	RIKYASKYQAQ-----SANYWKSIGNRGLARLDVIGRRAREAFADWIRKN-----	366
Db	445	RPAHGKTEQGLSPKSKPFATDGSAMSHKLEKADLL--KAIGEOILAN-VHSNDYFEVI	501
QY	367	--GKSANYGVGLSLEKAYIEGKAKNREMTYISETLFGSTEYVRRQAOFANALATYPAHA	424
Db	502	DFASDATITTRNGKVFYFADKDS-VTLPTQPOVEFLSGHVVRKE-----KPIONO	553
QY	425	GIKLSLDKX-----KDYLPST-DKRYLPAMLDIVRRIRPADKLPIDFKNVIDKK	473
Db	554	A--KSVYETVTOFTPLNPDDDPRLKDKTKLTKTI--AIGPTTSQELLAQAOSILNN	609
QY	474	EKGDTKKYADFVFDKSVYPYSDKFHAMLKSMDEKFEKAKAIEKDPAVELSKVYIAARAIO	533
Db	610	HPG-----YTIYERDSSIVTHDNDIFPTILPMDGEFYRVKKNNEQAYRINKSGLUEELNN	665
QY	534	ADAMANA-YAIEGK-----RLFFAGLREKYPECALPSPDANFT--MRMSYSTI	578
Db	666	TDLISEKYYVLKGEKPEYDFPRSHLKTITIKYVDVTNELKLSQOLLTASERINLDFRDL	725
QY	579	KGYEPDGA--WVN-----YHTTGKGVLEKQDPKSDFAVOENLILDFR--TKNYG	625
Db	726	--YDPRKAKALLNNLDARFIMDYTLTGK-----VEDNHDDTNRITTYVYG	769
QY	626	R--YAENGOLHIAF	637
Db	770	KRPEGENASVHLAY	783
RESULT 9			
AAW21727	AAW21727	standard; protein; 1181 AA.	
XX	AAW21727;		
AC	AAW21727;		
XX	01-OCT-1997 (first entry)		
DT			
DE	Streptokinase/maltose binding protein fusion protein, rskdelta14.		
DE			
XX	Plasminogen-binding fragment; streptokinase; degradation; MBP.		
KW	thrombolytic agent; blood clot; bolus; maltose-binding protein.		
KW			
XX	Streptococcus equisimilis.		
OS			
XX			
XX			
FH	Key	Location/Qualifiers	
FT	Protein	1..381	
FT		/label= Maltose_binding_Protein	
FT		/note= "acts as blocking group"	
FT	Protein	382..1181	
FT		/label= Modified_Streptokinase	

```

FT      /note= "Has N-terminal 14 amino acids deleted"
XX      WO9641883-A1.
XX      27-DEC-1996.
PD
XX      07-JUN-1996;    96WO-US09640.
XX      PF
XX      09-JUN-1995;    95US-0488940.
XX      PR
XX      (HARD ) HARVARD COLLEGE.
XX      PA
XX      Reed GL;
XX      PI
XX      WPI; 1997-065469/06.
XX      DR
XX      Modified forms of streptokinase resistant to enzymatic cleavage -
XX      PT      useul as thrombolytic agents in treating thrombosis and in medical
XX      PT      equipment
XX      PS
XX      Example 1; Page 12; 65pp; English.
CC      This sequence represents a fusion protein between maltose-binding
CC      protein and a mutant form of the the plasminogen-binding fragment
CC      of streptokinase which has the N-terminal 14 amino acids deleted.
CC      This fusion protein was used in the design of a modified streptokinase
CC      has an in vitro degradation rate at least 2 times slower than that of
CC      native streptokinase. Compounds containing modified streptokinases
CC      are specifically used as thrombolytic agents for dissolving blood
CC      clots in vivo in a mammal, preferably at a dose of 20000 U/Kg,
CC      optionally as a bolus rather than by continuous infusion.
CC      CC
CC      Sequence    1181 AA;
SQ

Query Match      3.3%; Score 122.5; DB 18; Length 1181;
Best Local Similarity 19.3%; Pred. No. 0.32;
Matches 130; Conservative 99; Mismatches 254; Indels 191; Gaps 35;

QY      75 GITVSDGGL-----IFNHHGCGAIGSQSFVHDHDLRDGFSVSRMGEELPIP---GLS 125
DB      558 GGLTTSQELLAQAOSILNKNHPGYTTERDSTI---YTHNDIFRTI---LPMDOETTYR 611
QY      126 VYLLRKIVKVTQKVGQGLGKGTDEBERLKAQEOCEIAXKB----- 167
DB      612 VKNRQAVRINK-----SGLNEELNNNDILSEKYYVLKKGKRPDPDRSHLKLFTIKY 666
QY      168 -NADENQLCIVEPEYSNNY--FLIVYDFKDVRAVFPASSVGKFG-----GDTDN 216
DB      667 VVDVFNELILKSEQLTLASERINDFPDLDPRDKALTY---NNLDAFGIMDTTLGKYED 723
QY      217 WMMPPHTGDFSFERYAAGADNRPAEYSKDNKKRYKRVYFAAVSMQGYKKADVMTIGFFGS 276
DB      724 -----NHDNTNRITTYVMG--KRP--EGENASIHLYA-----D 752
QY      277 TDRY-----LTSWGEDRIENENNRI-----EYRGIKQGIWKAEAMGADOAT 318
DB      753 KRRYEEEREVEYSYLRYCTGPIPDNPNKNNQGLVAVSAGYVEGINODISLKFEDIDLS 812
QY      319 RIKVASKYAQ-----SANYWKSIGMNGLARLDVIGKRRAEERAFADWIRKN----- 366
DB      813 RPAHGKTEQGLSPKSPFATDSGAMSHKLEKADLL--KAIQEQLIAN-VHSNDQFEVI 869
QY      367 --GKNAVYGDVLSLEKAKYKEGAKANREFTYISLFLFGTEVYRRARQAFANALATPDMAH 424
DB      870 DFASDATTTIDRNGKYAFADKDGs-VTLPTQPOEFLLSGHAVRKE-----KPIQO 921
QY      425 GILKSLDDKY-----KDYLFSL--DRKVLPAMLDIVRRRIADKLDPDIEKNAIDK 473
DB      922 A--KSDVDEYITQFLPMPDDDFRGGLADTKLKLTL--AIGGTTISQELLAQAOSILKN 977
QY      474 FKGDTKKTADEVEDKSVVPSDKFPAHLKSMDEKFAKAIKEDPAVELSKSVIAAARAIQ 533
DB      978 HGG-----YTIYERDSISYTHNDIDFRITLPMQOEFYVKKNEQYVRIKNGSGLNEEINN 1033

```

QY 534 ADAMANA-YAIEKGR-----RLEFAGLREMPGRALPSPDANFT---MRMSYCSI 578
 Db 1034 TDLISEKYYVLKKGKRPYDPDRSHLKLFTIKYVDVDTNELLSQQLLTASERNLDFRDL 1093
 QY 579 KGYEPQDGA--WYN-----YHTTGKGVLEKODPKSDEFVAOENILDLFR--TKNYG 625
 Db 1094 --YDPRDKAKLLYNLDAFGIMDYTLTGK-----VEDNHDDTNRITTYVMG 1137
 QY 626 R--YAENGOLHIAP 637
 Db 1138 KRPEGENASYHLAY 1151
 RESULT 10
 AAW21724
 ID AAW21724 standard; protein; 1194 AA.
 AC AAW21724;
 XX
 XX 01-OCT-1997 (first entry)
 DT
 XX Modified streptokinase, rSK5mut.
 DE
 XX plasminogen-binding fragment; streptokinase; degradation; MBP;
 KM thrombolytic agent; blood clot; bolus; maltose-binding protein.
 XX
 OS Streptococcus equisimilis.
 XX
 XX Synthetic.
 OS
 FH key Location/Qualifiers
 FT Protein 1..381
 FT /label= Maltose-binding-protein
 FT /note= "acts as blocking group"
 FT Protein 382..1194
 FT /label= Modified-streptokinase
 FT MISC-difference 391
 FT /label= R10A
 FT MISC-difference 417
 FT /label= R36A
 FT MISC-difference 426
 FT /label= R45A
 FT MISC-difference 432
 FT /label= R51A
 FT MISC-difference 440
 FT /label= R59A
 XX
 XX MO9641883-A1.
 PN
 XX 27-DEC-1996.
 PD
 XX 07-JUN-1996; 96MO-USO9640.
 PF
 XX 09-JUN-1995; 95US-0488940.
 PR
 XX (HARD) HARVARD COLLEGE.
 PA
 XX
 PI Reed GL;
 XX
 DR WPI; 1997-065469/06.
 XX
 PT Modified forms of streptokinase resistant to enzymatic cleavage -
 PT useful as thrombolytic agents in treating thrombosis and in medical
 PT equipment
 PT
 XX
 PS Claim 15; Page 41-44; 65pp; English.
 XX
 CC This sequence represents a fusion protein between maltose-binding
 CC protein and a modified form of the plasminogen-binding fragment of
 CC streptokinase containing 5 point mutations. This modified streptokinase
 CC has an in vitro degradation rate at least 2 times slower than that of
 CC native streptokinase. Compounds containing modified streptokinases
 CC are specifically used as thrombolytic agents for dissolving blood

CC clots in vivo in a mammal, preferably at a dose of 20000 U/kg,
 CC optionally as a bolus rather than by continuous infusion.
 XX
 SO Sequence 1194 AA;
 Query Match 3.3%; Score 122.5; DB 18; Length 1194;
 Best local similarity 19.3%; Pred. No. 0.33;
 Matches 130; Conservative 99; Mismatches 254; Indels 191; Caps 35;
 QY 75 GITVSDGL-----IFNHHGCGYGAISQSTVDDHDIADGCFVSTMGCELPIP---GLS 125
 Db 571 GDTTISQELLAQAOSILKNKHPGYTIYERDSSI--VTHNDIDFTI--LPMQDEFTYR 624
 QY 126 VKYLKIKYVTDKVEGOLKGITDEMERLKAQVEQELAKKE----- 167
 Db 625 VKNEQAVRIKK-----SGINEIINTDLISEKYYVLKKGKRPYDPDRSHLKLFTIKY 679
 QY 168 -MADENOLCIVEPFSYNNY--FLIYDYFKDYRMVFPAPSSVKGFG-----GDTDN 216
 Db 680 VDVPDTNELLSQQLLTASERNLDFRDLYDPRDKAKLLY--NNLDARGIMDYTLTGKVED 736
 QY 217 MMWRHHTGDFSVFRVYAGADNRPAEYSKDNKPYKPYFAAYSMGYKADVAMTIGFPGS 276
 Db 737 ----NHDDTNRITTYVMG--KR--EGENASYHLAY-----D 765
 QY 277 TDRY-----LTSWGVEDRIENENPRI-----EVRSIKGIMWEAMSADQAT 318
 Db 766 KDRYTEEREVEYSLRYTGTPIPDNPNKNSQLVSVAGVEGTNDISLKFELDTS 825
 QY 319 RIKYASKYAQ-----SANYKNSIGMNRGLARLDVIGRKRAERAPADWIRKN----- 366
 Db 826 RPAHGTKEQGLSPKSKPFATDSGAMSHKLEKADLL--KAIOEOLIAN-VHSNDYFEVI 882
 QY 367 --GKSAYGVDLSSLEKAYKEGAKANEMTYLSTLGGTFVEFAQFANALANPDAAH 424
 Db 883 DFASDATTITDNKGVYPADKGS--VTLPTQVQEFLLSGHVRVYKE-----KPIQNO 934
 QY 425 GILKSLDDKY-----KDYLPSTL-DRKVLPAMLDIYRRIRPADKLPIFFKNVIDKK 473
 Db 935 A--KSVDEYEVQFTPLNPDDDFRGLKDTKLTLT--AIGDTTISQELLAQAOSILKN 990
 QY 474 FKGDTKRYADEFEDKSVVPYSDKFNAMLKSMKDEKFAKALEKDPAVELSKSVIAARAIO 533
 Db 991 HPG--YTIYERDSSIYTHNDIFRTILPMDQETFRVKKRQAVYIRNKSGLNEIRIN 1046
 QY 534 ADAMANA-YAIEKGR-----RLEFAGLREMPGRALPSPDANFT---MRMSYCSI 578
 Db 1047 TDLISEKYYVLKKGKRPYDPDRSHLKLFTIKYVDVDTNELLSQQLLTASERNLDFRDL 1106
 QY 579 KGYEPQDGA--WYN-----YHTTGKGVLEKODPKSDEFVAOENILDLFR--TKNYG 625
 Db 1107 --YDPRDKAKLLYNLDAFGIMDYTLTGK-----VEDNHDDTNRITTYVMG 1150
 QY 626 R--YAENGOLHIAP 637
 Db 1151 KRPEGENASYHLAY 1164
 RESULT 11
 AAW21726
 ID AAW21726 standard; protein; 1194 AA.
 AC AAW21726;
 XX
 XX 01-OCT-1997 (first entry)
 DT
 XX Streptokinase/maltose binding protein fusion protein, rSK.
 DE
 XX plasminogen-binding fragment; streptokinase; degradation; MBP;
 KM thrombolytic agent; blood clot; bolus; maltose-binding protein.
 XX
 OS Streptococcus equisimilis.
 XX

QY	474	FKGDTKKIADPEVPEKSVIPYSDKEFHAMLSKMDKEKPAKALEKNDPAVELSKSVIAAARAQ	533
Db	991	HPG---YTVIEROSSIVTHDNDIFPTILPMDEFTYRVKNRQAVRIKKSGLMEIINN	1046
QY	534	ADANAAN-VAIIEKK-----RLFPAGLREMPGRALPSDANFT---MRMSYGS	578
Db	1047	TDLISEKYYVLAKGKERFYDFPDSHLKFTIKYVDVDTNELLSBQDLTAASEKNDLFRDL	1106
QY	579	KGYEPDGA--WYN-----YHTTGKVLKODPKSDPEFAVQENILDLFR--TKNYG	625
Db	1107	--YDPRKAKLTLNNDLAFGIMDYTLTGK-----VEDNHDDTRNIIITVYWG	1150
QY	626	R-YAENGQLHIAF	637
Db	1151	KRPEGENASYHLAY	1164
RESULT_12			
AAV34536			
ID	AAV34536	standard; Protein; 951 AA.	
XX	AC	AAV34536;	
XX	DT	25-AUG-1999 (first entry)	
XX	DE	Porphyromonas gingivalis protein PG67.	
XX	DE	Porphyromonas gingivalis; PG; periodontal disease; gingivitis;	
KW		vacine; antigenic.	
XX	OS	Porphyromonas gingivalis.	
XX	PN	WO9929870-A1.	
XX	PD	17-JUN-1999.	
XX	PF	10-DEC-1998; 98WO-AU01023.	
XX	PR	04-AUG-1998; 98AU-0005028.	
XX	PR	10-DEC-1997; 97AU-0000839.	
PR	31-DEC-1997;	97AU-0001182.	
PR	30-JAN-1998;	98AU-0001546.	
PR	10-MAR-1998;	98AU-0002264.	
PR	09-APR-1998;	98AU-0002911.	
PR	23-APR-1998;	98AU-0003128.	
PR	05-MAY-1998;	98AU-0003338.	
PR	22-MAY-1998;	98AU-0003654.	
PR	29-JUL-1998;	98AU-0004917.	
XX	PA	(CSLC-) CSL LTD.	
XX	PI	Agius CT, Barr IG, Hocking DM, Margetts MB, Patterson MA;	
XX	PI	Ross BC, Rothel LJ, Webb EA;	
XX	DR	WPI: 1999-385613/32.	
XX	DR	N-PSDB: AAX91754.	
XX	PT	Antigenic Porphyromonas gingivalis peptides for preventing	
XX	PT	gingivitis	
XX	PS	Claim 1; Page 526-527; 588pp; English.	
XX	PS	AAV34536 to AAV34536 encode two hundred and sixty six antigenic	
CC	CC	Porphyromonas gingivalis (PG) polypeptide sequences given in AAV34318 to	
CC	CC	AAV34483. AAV341802 to AAV34189 represent PCR primers used in the	
CC	CC	isolation of the PG polypeptides. The PG polypeptides have antibacterial	
CC	CC	activity with a vaccine mechanism of action. The PG polypeptides can be	
CC	CC	used as vaccines especially against Porphyromonas gingivalis. Probes can	
CC	CC	be used to detect Porphyromonas gingivalis in standard hybridisation	
CC	CC	assays. Porphyromonas gingivalis is involved in periodontal disease	
CC	CC	especially gingivitis.	

Seq	Sequence	951 AA;	3.2%;	Score 117.5;	DB 20;	Length 951;
Query Match	Best Local Similarity	20.3%;	Pred. No. 0.63;			
Matches	95;	Conservative	65;	Mismatches	190;	Indels 119; Gaps 22;
QY	248	PKKPYFAVAVSOGKAKADYAMTIGFPGSTDBYVLWSWG---- <td>302</td> <td></td> <td></td> <td></td>	302			
Db	263	PSNKRYSGGILMPTYGEDN--RIGF-----YLNKGGYFAFSQID-----LALRGE	307			
QY	303	-IKQIKWEAMSADQATRIKYASKAQSANWYKNSIGNMRGLARLDVIGRKAEREAFAD	361			
Db	308	IFSKKSWGIGSASQSKKKRYKNGSF--EANY-----	336			
QY	362	WIRKNGKSAVYGDVLS---SLEKAYKEGAKANREMYLSETLFGTEVYRFAQFANALA	417			
Db	337	LVSKSGDKYVPGDYSKTSLNIRWTHSODPKANP-----LQTLISANVNPAATSYFONSIN	391			
QY	418	TNPDAHA--GILKSLDDKVKDYLPSLDRKVLPLAMLDI--VRRRIPADKLPDIFKVIDX	472			
Db	392	TTYDVNATPATTTSSAVSISRKFPGTPPSITGSM-DISQNMDDTYSVLTPLNLSIMSTR	450			
QY	473	-KFKGDTKKYADFVDSVYPYSDFKFAMLKMDKEKFALEKDPAYELSKVIAARA	531			
Db	451	YPFKKTKTVGPRWRYEKLSVSGQLRNSILTKKEDLLOSNLVDRKNGMRHSVPISLTV	510			
QY	532	IQADAMANAVALAEKGRLEFAGLRMY--PGRALPSDANFTMRMSYSGISIKYEPDQDAM	588			
Db	511	PLDDYINILTMGVNVMEMWYTKGIRKSMNDDKKTPLPSDTTYKFRRLY---DYSLSAG--	564			
QY	589	YNYHTTGCGVLEKQDPKSDFEFAVQENILD-----FRKINYGRIY-----	628			
Db	565	--LSTTLTGMEFPMKP---FSFGGNLIMIRHRTPTVSFSYMPFTKRRYGFWELEHT	618			
QY	629	-ENGOLHIAFLSN--NDITG---GNSGSPYDXKNGRLGLGAFDQWMEA	670			
Db	619	DONGRLHTLTSYPEREQIFGAPSMGNAGS-----VNESFDNNLEA	658			
RESULT 13						
ID	AAI34403					
ID	AAI34403	standard; Protein; 953 AA.				
XX	AAI34403;					
XX	25-AUG-1999	(first entry)				
DT	DE	Porphyromonas gingivalls protein PG67.				
XX	XX					
KW	XX	Porphyromonas gingivalls; PG; periodontal disease; gingivitis;				
KW	XX	vaccine; antigenic.				
XX	OS	Porphyromonas gingivalls.				
XX	FN	W09929670-A1.				
XX	PD	17-JUN-1999.				
XX	PF	10-DEC-1998; 98WO-AU01023.				
XX	PR	04-AUG-1998; 98AU-0005028.				
XX	PR	10-DEC-1997; 97AU-0000839.				
XX	PR	31-DEC-1997; 97AU-0001182.				
XX	PR	30-JAN-1998; 98AU-0001546.				
XX	PR	10-MAR-1998; 98AU-0002264.				
XX	PR	09-APR-1998; 98AU-0002911.				
XX	PR	23-APR-1998; 98AU-0003128.				
XX	PR	05-MAY-1998; 98AU-0003338.				
XX	PR	22-MAY-1998; 98AU-0003654.				
XX	PR	29-JUL-1998; 98AU-0004917.				
XX	PA	(CSLC-) CSL LTD.				

Query Match	3.2%	Score 117.5	DB 20	Length 953
Best Local Similarity	20.3%	Pred. No. 0.63		
Matches	95	Conservative	65	Mismatches 190; Indels 119; Gaps 22
0Y	248	PKPYVFAVSAQGRKADDDYATMTGPPGSDTDYLTSMG---- <td>302</td> <td></td>	302	
Db	265	PSKNTYSSGILMPTYGEDN---RYGF-----YLRNGGYFAFSDYID-----LALRGE	309	
0Y	303	-IKGIIMKEAMSADQATRIKAYSKYAQSANWKNISGMNGLARLDYIGKRAERAFAD	361	
Db	310	IFSQSGWGISAQSKYKKRYKNGSF--EANY-----	338	
0Y	362	WIRKNGKSAVYGDVLS----SLEKAKBEAKANREMTYISLTLFGCTEYVYRQAFANALA	417	
Db	339	LVSKSGDKYVPDYSKTTSLNIRWTHSDPKANP-----LQTLISANVNPAFGSYEQNSLN	393	
0Y	418	TMPDAHA--GILKSLDDKDYLPISDRKRVLPAMLDI---VRRRIPADKLPIQFNKVIDK	472	
Db	394	TYIDVNAKRAITRRSSAVSISKRFPGPFSITGSM-DISQNMDDTYSLLPLNLSIMSTR	452	
0Y	473	-KFKGDYTKKADVFEDKSVVPSDKFHAMLKSMDEKFKAKIEKDAVELSKSVIAAABA	531	
Db	453	YFKKRTVRGPERMYEKILSVGSGOLRNSILTKERKDLQSNLIVRWKNGMRHSVPISLTIV	512	
0Y	532	IGADAMANAVALTEKGRLEFFAGLREMY---PGRALPSDANFTMKMSYGIKGYEPQDGMW	588	
Db	513	PLLDVYINILTMGVNVMWYTKGIRKSWNEDKTKTFLPSDTTYKFRRLY---DYLSAG--	566	
0Y	589	YVNHHTTGKGVLEKOPKSDPEFAVQENIIDL-----FRTKNYGRYA-----	628	
Db	567	--LSTTLTYMEFPKWP---FSFGGLIMIRKRRFTYVSFSYMPDPTRKRYGFELLEHT	620	
0Y	629	-ENGOLHIAFLSN--NDING---GNSGSPVEDKNGRLIGLAFDGNMEA	670	
Db	621	DQNGKHLTLTLYSPYFEQIGAPSGMAGS-----VNFSPDNNLEA	660	
RESULT 14				
AA88649				
ID	AA88649	standard; Protein; 711 AA.		
XX	AA88649;			
XX	02-SEP-1996	(first entry)		
XX	DE	Neisseria meningitidis B163 transferrin receptor Tbp2 subunit.		
XX	KM	Transferrin receptor; Tbp2 subunit; deletion mutant; vaccine;		
XX	KW	passive immunisation; immunotherapy; IM2169; IM2394.		
XX				

```

OS Neisseria meningitidis (strain B2163).
XX
FH Key Location/Qualifiers
FT Protein 1..711 /label= Tbp2_subunit
FT Peptide 1..20 /label= signal_peptide
FT Domain 21..366 /label= Domain_1
FT /note= "N-terminal domain"
FT Domain 367..564 /label= Domain_2
FT /note= "hinge domain"
FT Domain 565..711 /label= Domain_3
FT /note= "C-terminal domain"
XX
PN WO953049-A2.
PD 07-DEC-1995.
XX
PF 30-MAY-1995; 95MO-FR00701.
XX
PR 31-MAY-1994; 94FR-0006594.
XX
PA (INMR ) PASTEUR MERIEUX SERUMS & VACCINS.
PA (TRGE ) TRANSGENE SA.
XX
PI Jacobs E, Legrain M, Mazarin V, Lissolo L;
PI Millet MBJ;
XX
DR MPI: 1996-030562/03.
DR N-PSDB; AAT11244.
XX
PT Polypeptide(s) for vaccination against Neisseria meningitidis group
PT B - comprising deletion mutants of transferrin receptor Tbp2
PT subunit
XX
PS Disclosure: Page 82-87; 114pp; French.
XX
CC The present sequence is that of the N.meningitidis strain B2163
CC transferrin Tbp2 subunit. The Tbp2 polypeptide has three
CC domains (see features table); deletion mutants in which at least
CC one of the domains is partially or totally deleted are claimed,
CC provided that the first and second domains are not simultaneously
CC partially or totally deleted. The positions of the 3 domains in
CC B2163 are defined by alignment with the IM2169 sequence. The
CC deletion mutant polypeptides of the invention can generate an
CC immune response against N.meningitidis.
XX
SO Sequence 711 AA:
Query Match 3.1%; Score 114; DB 17; Length 711;
Best Local Similarity 18.7%; Pred. No. 0.81;
Matches 134; Conservative 97; Mismatches 264; Indels 220; Gaps 34;

```

```

Db 254 KHEGYFTSLLEVDFSGSKLTKGLIRNNRYTNATNDKTYTQYSDAQITGNRNGKAI 313
Oy 332 -----YKNSTIGANRGL-----ARLDVIGKRAEEPAFD 361
Db 314 ATDKPTGTGTLKLPFVSDSSLSGFGFPGKEELGFRFLSDDKVAVVGSATKTKDYT--- 370
Oy 362 WIRKNGKSAVGDVLSLEKAYKEGAKANREMYLS--ETLPGCTEVRRAOFANL--- 416
Db 371 ---ENGAVASGTDAAASNGAGTSSNSKLTIVLAVELKLDQKEVQKIDNFSNAAQLV 427
Oy 417 -----ATNPAAHAGILKSLDDKYK-DYLPISLDRKVLPLAMLDIVRRRIIPADK 461
Db 428 VDCIMIPILPETSESGNQNAGTNGTAVTRKFDHTPESDKKDAQAGTQTNAGTASNT 487
Oy 462 LPDIFKAVIDKKKGTCKKADVFEDKSVYPYDKRHAMLSKDKKEPAKIKDPAVEL 521
Db 488 AGDT-----NGKTKY-----EVEVCCSNLYTKYGLTRKN----- 519
Oy 522 SKSVIAARAIOADAMANAIAIEKGRLPFAGLREYVPGRLPSDANFTWMSYSGIKY 581
Db 520 SKS---AMQGESSQADATBEVQGSMLQG--ERTDEKEIRSEQNIYR----- 555
Oy 582 EPQDGAWYNY--HTGKG-VLEKODPKSDFAVOENIIDFTKXVGRYAENGOLHAF 637
Db 566 ---GSWGHIASTSMGASDKEGNRAEFV-----NPEKKIKGTI---- 606
Oy 638 LSNNDITGNGSPVPEDKNGRLIGLAFDGNWEMSGDIEPEPLQRTISVDIRY 692
Db 607 -----TAENROEATFTIDKIEGNGFSGT--AKTAEIGFDLDQKNYTRPKAYI 653

RESULT 15
AAB43567
ID AAB43567 standard; Protein; 708 AA.
AC AAB43567;
DT 08-FEB-2001 (first entry)
XX
DE Human cancer associated protein sequence SEQ ID NO:1012.
XX
KW Human; cancer associated gene; cancer antigen; detection; cancer;
KW diagnosis; cytostatic; proliferative; vulnerability; immunomodulator;
KW antidiabetic; antiasthmatic; antirheumatic; antiarthritic; antiviral;
KW antiinflammatory; antithyroid; antiallergic; antibacterial; cardiac;
KW dermatological; neuroprotective; thrombolytic; coagulant; noctropic;
KW vasotropic; antipsoriatic; antiangiogenic; gene therapy; inflammation;
KW immune disorder; haematopoietic cell disorder; autoimmune disorder;
KW allergic reaction; graft versus host disease; organ rejection;
KW haemostatic; thrombolytic; cardiovascular disorder; infection;
KW neurological disease; drug screening.
XX
OS Homo sapiens.
XX
PN WO200055350-A1.
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US05882.
XX
PR 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR MPI: 2000-587533/55.
XX
DR N-PSDB; AAC77776.
XX
PT Novel isolated nucleic acids comprising sequences encoding peptides
XX useful for treating or diagnosing e.g. cancer -

```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 17, 2002, 22:36:01 : Search time 24 Seconds
(without alignments)
724.626 Million cell updates/sec

Title: US-10-008-355-2

Perfect score: 3719

Sequence: 1 MWMKLKSLILGALLLGASG.....LFMDXWGCPRILQELKLI 712

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Maximum Match 0%

Listing first 45 summaries

Database :

Issued Patents_AA:*
1: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/prodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/prodata/2/1aa/6A_COMB.pep:*
4: /cgn2_6/prodata/2/1aa/6B_COMB.pep:*
5: /cgn2_6/prodata/2/1aa/PCBUS_COMB.pep:*
6: /cgn2_6/prodata/2/1aa/Backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	130	3.5	883	2	US-08-953-492-2
2	127.5	3.4	1194	2	US-08-488-940-18
3	122.5	3.3	800	2	US-08-488-940-4
4	122.5	3.3	813	2	US-08-488-940-3
5	122.5	3.3	1181	2	US-08-488-940-2
6	122.5	3.3	1194	2	US-08-488-940-1
7	122.5	3.3	1194	2	US-08-488-940-17
8	110	3.0	1027	4	US-08-446-137B-2
9	109	2.9	1577	2	US-08-793-824-2
10	106	2.9	972	3	US-08-335-844A-23
11	103.5	2.8	1114	4	US-08-811-583-2
12	102.5	2.8	833	4	US-09-514-302-3
13	102.5	2.8	1938	4	US-09-514-302-2
14	101.5	2.7	747	1	US-07-854-596B-40
15	101	2.7	433	1	US-08-417-492-2
16	100.5	2.7	638	1	US-08-712-241-6
17	100.5	2.7	885	1	US-08-484-105-14
18	100.5	2.7	885	1	US-08-484-106-14
19	100	2.7	15281	2	US-08-471-119A-2
20	99.5	2.7	722	4	US-08-961-083-84
21	99	2.7	725	2	US-08-816-105A-1
22	99	2.7	908	2	US-08-349-380-2
23	98.5	2.6	637	1	US-08-712-241-5
24	98.5	2.6	637	4	US-08-026-143B-4
25	98.5	2.6	637	5	PCT-US92-10621-4
26	98.5	2.6	637	5	PCT-US94-02233-4
27	97.5	2.6	468	4	US-08-961-083-10

28	97.5	2.6	1430	3	US-09-008-172-2	Sequence 2, Appl1
29	97.5	2.6	1430	4	US-09-210-361-6	Sequence 6, Appl1
30	97	2.6	890	4	US-09-342-648-10	Sequence 10, Appl1
31	96.5	2.6	654	1	US-08-487-890A-12	Sequence 12, Appl1
32	96.5	2.6	654	2	US-08-478-435-12	Sequence 12, Appl1
33	96.5	2.6	654	2	US-08-337-483-12	Sequence 12, Appl1
34	96.5	2.6	654	2	US-08-478-373-12	Sequence 12, Appl1
35	96.5	2.6	654	3	US-08-474-671-12	Sequence 12, Appl1
36	96.5	2.6	654	3	US-08-483-577A-12	Sequence 12, Appl1
37	96.5	2.6	654	4	US-08-897-438-12	Sequence 12, Appl1
38	96.5	2.6	654	4	US-08-637-654-12	Sequence 12, Appl1
39	96.5	2.6	1128	1	US-08-111-939-2	Sequence 2, Appl1
40	96.5	2.6	1181	1	US-08-053-614-4	Sequence 4, Appl1
41	96.5	2.6	1181	1	US-08-316-397B-4	Sequence 4, Appl1
42	96.5	2.6	1181	2	US-09-034-306-4	Sequence 4, Appl1
43	96.5	2.6	1181	4	US-09-259-437-4	Sequence 4, Appl1
44	96.5	2.6	1181	5	PCT-US93-09782-4	Sequence 4, Appl1
45	96	2.6	1007	4	US-08-961-083-216	Sequence 216, App

ALIGNMENTS

RESULT 1
US-08-953-492-2
Sequence 2, Application US/08953492
Patent No. 5849555
GENERAL INFORMATION:
APPLICANT: Brown, James
APPLICANT: Jaworski, Deborah
APPLICANT: Lawlor, Elizabeth
APPLICANT: Wang, Min
TITLE OF INVENTION: NOVEL VALS
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESSES:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/953,492
FILING DATE: 17-OCT-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/844,064
FILING DATE: 18-APR-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9607991.8
FILING DATE: 18-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Gimm1, Edward R
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P31458-4/1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 883 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-953-492-2

Query Match	3.5%;	Score 130;	DB 2;	Length 883;
Best Local Similarity	20.4%;	Pred. No. 0.0042;		
Matches 104;	Conservative 63;	Mismatches 168;	Indels 176;	Gaps 24

```

QY 192 DVF-----KQVRNMFAPSSSGCGSDTMMNRHNGDSEVFVYAGADNRPREYS 243
Db 26 DVERPSGDQKKRPSIYIPPNWTKL-----HLG-----HA 57
QY 244 KDNRPKYRPUYEAAYSMOGYKADYAMTIGFPGSTDRLY-TSWGEDRIENENNPRIEVRG 3024
Db 58 WDT-TLQDIIIRKRMGOFD-----TLMJPGMHAGIATQAKVEERLREGEISRYDLGR 110
QY 303 -----IKGCIWKEAMSADQATPIKA-SKYQASANYWKNISGMRGLARLDYIGRKRAEER 357
Db 111 ESPLTKWEMKD-----EATATIKQOMGKMGISLVDSYSERFTLEGS-----KAVRK 158
QY 358 AFADMIK-----NGKSAVYGD-----VLSSLEKAYKSGARANEMTYL----- 3966
Db 159 VFVDLYKKGMWYIRGEFIJINMDPARFATLSDEIYIHKVBECAFYHNMNMLDGSRVLEVAT 218
QY 397 --SETLEGGTEVYVFAOPANALATNPANAHGILKSLDDTKDYILPS-----LDRKYLPAH 449
Db 219 TRPETMEGDV-----AVALNPE-----DPRKDLIGKNVILPIANKILPIY 259
QY 450 LD-----IYRRIRPADKLTDIEKNYIDKKFKSDTKKYADVFEDKSUVPRYSDK 4966
Db 260 GDEHADPEFGGVYKTIIPAHDPNDFLVGGQHNLPRQVVMYMDDGTMMNDLAREFGM---DR 316
QY 497 FHAMLSMDKKEFKPALEKDPAYELSSVYLAARA-----IQAD 535
Db 317 FEAR-----KAVVAKLEIGALVIEKRVSHVGHSEFTGVVVERISTONFVKMDOLAKN 371
QY 536 AMANAYALEKKRLEFFAGLREMYGKALPS-----DANFTMRYSYSGIKGYEPDQ 5866
Db 372 AIANQDTEDK-----VEFYRRPNDFLOWMENVMHWIVSRQJMWCH-----QIP 416
QY 587 AMYNHTTGKGVLEKODPKSDEFAVOENILD 617
Db 417 AMYN--ADGEMYVEEAPREGGGMVQDDVDLD 445

```

RESULT 2
 US-08-488-940-18
 : Sequence 18, Application US/08488940
 : Patent No. 5854049
 : GENERAL INFORMATION:
 : APPLICANT: Reed, Guy L.
 : TITLE OF INVENTION: PLASMIN-RESISTANT STREPTOKINASE
 : NUMBER OF SEQUENCES: 20
 : CORRESPONDENCE ADDRESSES:
 : ADDRESSEE: Fish & Richardson P.C.
 : STREET: 225 Franklin Street
 : CITY: Boston
 : STATE: MA
 : COUNTRY: USA
 : ZIP: 02110-2804
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: PatentIn Release #1.0, Version #1.30
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/488,940
 : FILING DATE: 09-JUN-1995
 : CLASSIFICATION: 514
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Fraser, Janis K.
 : REGISTRATION NUMBER: 34,819
 : REFERENCE/DOCKET NUMBER: 05433/009001
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: 617/542-5070
 : TELEFAX: 617/542-8906
 : TELEX: 200154

```

; INFORMATION FOR SEQ ID NO: 18:
;
; SEQUENCE CHARACTERISTICS:
;     LENGTH: 1194 amino acids
;     TYPE: amino acid
;     STRANDEDNESS: not relevant
;     TOPOLOGY: linear
; MOLECULE TYPE: protein
;
US-08-488-940-18

```

Query Match	3.4%	Score	127.5	DB	2	Length	1194
Best Local Similarity	20.1%	Pred. NC	0.012				
Matches	133	Conservative	99	Mismatches	265	Indels	167
						Gaps	35

```

0Y 75 GITYSDGL-----ITNHCYGAIQOSITYDHYLDADGVSTKMGELPIP-----GLS 125
Db 571 GDTTISOELLAQOSILINKNHPGTTIYERDSI--VTHDNDIFPRI--LPMOEFYTR 624
0Y 126 VKYLARKIYKVDKVEGQJAKITDEMERLTKAOCVCOELAKKE-----167
Db 625 VKNNGOAVRIKK-----SGLNEEINTDILISKYYVLKGGKRPYDPDRSLKLTITKY 679
0Y 168 -MADENOICLIVEPYSNNEY---FLIYVDYFKDVRNRYFAPSSVGKFG-----GDTDN 216
Db 680 VDVITNELKSEQLTISEENLDFRLOYDFRDKAKLY--NNLDAFGMDYTLGKVED 736
0Y 217 MWMPHRTGDSFVFRVYGAADNPAEYSKONKPYKPYFAAVMOGYKKAADYAMTIGFPGS 276
Db 737 ----NHDDTNIIIVYMG--KR--BGEASH-----LAYDADRYTEEBREYVS 778
0Y 277 TDRYLTSGNVEDRIENENNPRI-----EYKTIKQIGTKMEAMSADOATRKYASKYAO-- 328
Db 779 YLRY-TGTPIDNDNDKNNQOLVYVAGVTEGNODISLKFIEDLTSRPAHGGKTEQGL 837
0Y 329 ---SANYKNSIGNRGLARIDVYIGKRAAEARAFADMIKN-----GKSAYYGDVL 376
Db 838 SPKSKPRFATDSGAMSHKLEKADLL-KAIOEOJLAN-VHSNDYFEVIDFASDATITDRN 894
0Y 377 SLEKAIYEGAKANRENTYISETLFGTEYVYRPAOFANALATNPDAHAGILSKLDOK-- 434
Db 895 GKVFYADKDGs-VYLPYQPOEFLTSGHVARRYKE-----KPIONOA--KSDVVEYTV 944
0Y 435 -----KDLPLST--ORVYLPAMLDIVRRRIADKLPDIFKNVIDDKFKFGDTKRYAOFV 485
Db 945 OPTPLNPDDPRBLKOTKLITKL--AIGTTISOELLAQOOSILINKNHPG---YTYIE 998
0Y 486 FDKSVYVPSDKFHMLKSMDEKERFAAIEKDPAAVELSKSVIAAARAQADAMANA-VAIE 544
Db 999 RDSISIVHDNDIFPTLIPMOEFTYRYKKNQOAYRINKKSGLNEEINTDILSEKYYVLK 1058
0Y 545 GKG-----RLFFAGLREMYGRLALPSANPT--MMMSYGSIKGYEPDOGA--W 588
Db 1059 KGEKPYDPFDRSHKLTITTYVVDITNELKSSQILTASERNLDPRDL--YDPRKAKALL 1116
0Y 589 YN-----YHRTGKVLKEDOKPSDEFAYOENITLDFR--TNNYGR--YAENGOLHI 635
Db 1117 YNNIDAFGIMDYTLTKG-----VEDNHDDTNRILITTYMGKRPEGENASVHL 1162
0Y 636 AP 637
Db 1163 AY 1164

```

RESULT 3
US-08-488-940-4
; Sequence 4, Application US/08488940
; Patent No. 5854049
;
; GENERAL INFORMATION:
;
; APPLICANT: Reed, Guy L.
; TITLE OF INVENTION: PLASMIN-RESISTANT STREPTOKINASE
; NUMBER OF SEQUENCES: 20
;
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
;


```

? CITY: Boston
? STATE: MA
? COUNTRY: USA
? ZIP: 02110-2804
?
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.30
?
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/488,940
? FILING DATE: 09-JUN-1995
? CLASSIFICATION: 514
? ATTORNEY/AGENT INFORMATION:
? NAME: Fraser, Janis K.
? REGISTRATION NUMBER: 34,819
? REFERENCE/DOCKET NUMBER: 05433/009001
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 617/542-5070
? TELEFAX: 617/542-8906
? TELEX: 200154
?
? INFORMATION FOR SEQ. ID NO: 4:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 800 amino acids
? TYPE: amino acid
? STRANDEDNESS: not relevant
? TOPOLOGY: linear
? MOLECULE TYPE: protein
?
? US-08-488-940-4

```

Query Match 3.3%; Score 122.5; DB 2; Length 800;

Best Local Similarity 19.3%; Pred. No. 0.018;

Matches 130; Conservative 99; Mismatches 254; Indels 191; Gaps 35;

```

QY 75 GIVSOGU-----IFTNHCGYGAIOGOSTVDHDLRDGFVSRMGELPIP---GLS 125
DB 177 GDTTSEELAAQOSILNKHPGTYIERDSSI---VTHNDIFRTI---LPMOQETTYR 230
QY 126 VKYLRKIVKTVDRVEGLKGTDEMELRKAQOYCOELAKE-----167
DB 231 VKNREQAYRINKK---SGLNEEINNTDLISEKYVYLKKGKRPYDFDRSHLKLFTIKY 285
QY 168 -NADENOLCIVEPFYSSNEY---FLIYDYVFKVRYAFAPSSVGKFG-----GDTDN 216
DB 286 VDVDTNELLKSEQLTASEENLDPRDLYDPRDAKILY---NNLDAFGIMDYTLTGVED 342
QY 217 WMPRHTGDSVFRVYAGADNRPAYSKONKPYKPYFAAVSMQGYKADYAMTIGPPGS 276
DB 343 ---NHDDTRITTVYMG---KRP---EGENASYHLAY-----D 371
QY 277 TDRY-----LTSNGVEDRIENENNPR-----EVGIGIKGIMKEAMSADQAT 318
DB 372 KDRTEEREVEYSYLRGTGPIDNPNDKNNNSOLVSVAGTVEGTNDISLKEFEIDLTS 431
QY 319 RIKYASKYAO-----SANYKNSIGNMRGLARLDVIGRKRAREAFADWIRKN-----366
DB 432 RPAHGGTIEGSLSPKSPKFTIDSGAMSHKLEKADL--KAIOBELIAN--VHSNDYREVI 488
QY 367 --GKSAVYGVLSLEKAYEGAKANREMYLTSETLFGTEVVVFAQAFANALATNPDAHA 424
DB 489 DFASDATITTRNGKVFADKGS--VTLPTQVQEFLLSGHVRVYKKE-----KPIQNG 540
QY 425 GILKSLDDKY-----KDYLPST--DRKVLPRAMLDIYRRIRPAKLDPIFRANVDK 473
DB 541 A--KSVVETTVQFTPLNPDDDFPGKDKRLKLT--AGDITTSOELLAQOSILNK 596
QY 474 FKGDTRKYADFPDQSVYPSDKFHAMLSMDKEKFAKATEKDPFAVLSKSVIAARAIO 533
DB 597 HPG---YTIYERDSSIVTNDIFRTILPMDQETTYRVKNRREGAYVINKSGINEINN 652
QY 534 ADAMANA-YAIEKCK-----RLFAGLRKEMTPGALPSDANFT--MRMSYGST 578
DB 653 TDLISEKYVYLKKGKRPYDFDRSHLKLFTIKYVDVTNELKSEQLLTASERNLDFRDL 712

```

```

QY 579 KGYEPOGA--WYN-----YHTGKGVLEKQDPKSDFAVOENIDLFR--TKNYG 625
DB 713 --YDPRDAKALYNNMDARGIMDYTLTGK-----VEDNHDDTRITTVYMG 756
QY 626 R--YAENGOLHIAF 637
DB 757 KREGENASYHLAY 770

```

RESULT 4

```

US-08-488-940-3
? Sequence 3, Application US/08488940
? Patent No. 5854049
?
? GENERAL INFORMATION:
? APPLICANT: Reed, Guy L.
? TITLE OF INVENTION: PLASMIN-RESISTANT STREPTOKINASE
? NUMBER OF SEQUENCES: 20
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Fish & Richardson P.C.
? STREET: 225 Franklin Street
? CITY: Boston
? STATE: MA
? COUNTRY: USA
? ZIP: 02110-2804
?
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/488,940
? FILING DATE: 09-JUN-1995
? CLASSIFICATION: 514
? ATTORNEY/AGENT INFORMATION:
? NAME: Fraser, Janis K.
? REGISTRATION NUMBER: 34,819
? REFERENCE/DOCKET NUMBER: 05433/009001
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 617/542-5070
? TELEFAX: 617/542-8906
? TELEX: 200154
?
? INFORMATION FOR SEQ. ID NO: 3:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 813 amino acids
? TYPE: amino acid
? STRANDEDNESS: not relevant
? TOPOLOGY: linear
? MOLECULE TYPE: protein
?
? US-08-488-940-3

```

Query Match 3.3%; Score 122.5; DB 2; Length 813;

Best Local Similarity 19.3%; Pred. No. 0.019;

Matches 130; Conservative 99; Mismatches 254; Indels 191; Gaps 35;

```

QY 75 GIVSOGU-----IFTNHCGYGAIOGOSTVDHDLRGFVSRMGELPIP---GLS 125
DB 190 GDTTSEELAAQOSILNKHPGTYIERDSSI---VTHNDIFRTI---LPMOQETTYR 243
QY 126 VKYLRKIVKTVDRVEGLKGTDEMELRKAQOYCOELAKE-----167
DB 244 VKNREQAYRINKK---SGLNEEINNTDLISEKYVYLKKGKRPYDFDRSHLKLFTIKY 298
QY 168 -NADENOLCIVEPFYSSNEY---FLIYDYVFKVRYAFAPSSVGKFG-----GDTDN 216
DB 299 VDVDTNELLKSEQLTASEENLDPRDLYDPRDAKILY---NNLDAFGIMDYTLTGVED 355
QY 217 WMPRHTGDSVFRVYAGADNRPAYSKONKPYKPYFAAVSMQGYKADYAMTIGPPGS 276
DB 356 ---NHDDTRITTVYMG---KRP---EGENASYHLAY-----D 384
QY 277 TDRY-----LTSNGVEDRIENENNPR-----EVGIGIKGIMKEAMSADQAT 318

```


APPLICATION NUMBER: US/08/488,940
FILING DATE: 09-JUN-1995
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Fraser, Janis K.
REGISTRATION NUMBER: 34,819
REFERENCE/DOCKET NUMBER: 05433/009001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1194 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-488-940-1

Query Match 3.3%; Score 122.5; DB 2; Length 1194;

Best Local Similarity 19.3%; Pred. No. 0.035;
Matches 130; Conservative 99; Mismatches 254; Indels 191; Gaps 35;

QY 75 GITVSDQGL-----IFTNHCYGALIGOSQTVVDHDLRDGFVSRTMGELPIP---GLS 125
DB 571 GDTITSQELLAQAOSILNKNHPGTYIERDSSI---VTHDNDIFRTI---LPMDOEFYR 624
QY 126 VYLRKIVAVTDKVEGQLGTDMEERLKAQEVCOELAKKE----- 167
DB 625 VKNREGAVYRINKK-----SGLNEEINNTDLISEKYYVLKKGEPYDPFRSHLKLFTIKY 679
QY 168 -NADENQOLIVEPFYSNNY---FLIVYDFKDYRMVFPAPSSVGF-----GDTDN 216
DB 680 VVDVTNELLSQGLASRNLDPRDYLDPKAKLLY---NNLDAFGIMDTLITGKVED 736
QY 217 WMPRHTGDSVFYRYAGADNRPAYSKDNKPKPYFAAVSMOGYKADYAMTIGFPGS 276
DB 737 ----NHDDTNRIITYMG---KRP---EGENASYHLAY-----D 765
QY 277 TDYR-----LTSMGVEDRIENENPRI-----EVRGIGQIMKEAMSDAQAT 318
DB 766 KRYTEEREVYSYLRYTGTPIPDNDNKNNSQLVSVAGTVEGTDLSLKFELDLS 825
QY 319 RIKYASKYAQ-----SANWKNSIGMNRGLARLDVIGRRABERAFADIRKN----- 366
DB 826 RPAHGKTEQGLSPKSPKPRATDSGAMSHKLEKADLL--KAIDQOLIAN--VHSDDYFEVI 882
QY 367 --GKSAVYGDVLSLEKAYKEGAKANREMTYLSSETLFGTEVYRFAQFANALATNPDAHA 424
DB 883 DRASDATITDRNGKYVFADKDS--VTLPTQPOVQEFILSGHVRYKE-----KPIQNO 934
QY 425 GLIKSLDDXY-----KDYLSL--DRKVLPAMLDIVRRIRIPADKLPIEFKNVIDRK 473
DB 935 A--KSVADVEYVQETPLNDDDFRPLKDKTKLKTLL--AIGDTITSQELLAQAOSILNKN 990
QY 474 FKGDPFKYADFYVDKSVVPYSDFKHAMLSMOKEFAKAIEKDPAVELSKSVYAAARAQ 533
DB 991 HPG---YTIYERDSSIVTHDNDIFRTILPMDQEFYRYKKNRQOAVRINKKSGLINEINN 1046
QY 534 ADAMANA-YAIEKSK-----RLFFAGLREMYPGRALPSDANFT---MRMSYGS 578
DB 1047 TDLISKYVVLKGEKPYDPFRSHLKLFTIKYVVDVTNELLSQGLASRNLDPRDYLDPKAKLLY 1106
QY 579 KGYEPODGA--WYN-----YHTTGKGVLEKODPKSDEFAVQENIIDLFR--TKNYG 625
DB 1107 --YDPPDKAKLLYNNLDARGIMDYTLTGK-----VEDNHDDTNRIITYVYG 1150
QY 626 R--YAENGQLHTAF 637
DB 1151 KRPEGENASYHLAY 1164

RESULT 7
US-08-488-940-17
Sequence 17, Application US/08488940
Patent No. 5854049

GENERAL INFORMATION:
APPLICANT: Reed, Guy L.
TITLE OF INVENTION: PLASMIN-RESISTANT STREPTOKINASE
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/488,940
FILING DATE: 09-JUN-1995

CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Fraser, Janis K.
REGISTRATION NUMBER: 34,819
REFERENCE/DOCKET NUMBER: 05433/009001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154

INFORMATION FOR SEQ ID NO: 17:

SEQUENCE CHARACTERISTICS:
LENGTH: 1194 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-488-940-17

Query Match 3.3%; Score 122.5; DB 2; Length 1194;
Best Local Similarity 19.3%; Pred. No. 0.035;
Matches 130; Conservative 99; Mismatches 254; Indels 191; Gaps 35;

QY 75 GITVSDQGL-----IFTNHCYGALIGOSQTVVDHDLRDGFVSRTMGELPIP---GLS 125
DB 571 GDTITSQELLAQAOSILNKNHPGTYIERDSSI---VTHDNDIFRTI---LPMDOEFYR 624
QY 126 VYLRKIVAVTDKVEGQLGTDMEERLKAQEVCOELAKKE----- 167
DB 625 VKNREGAVYRINKK-----SGLNEEINNTDLISEKYYVLKKGEPYDPFRSHLKLFTIKY 679
QY 168 -NADENQOLIVEPFYSNNY---FLIVYDFKDYRMVFPAPSSVGF-----GDTDN 216
DB 680 VVDVTNELLSQGLASRNLDPRDYLDPKAKLLY---NNLDAFGIMDTLITGKVED 736
QY 217 WMPRHTGDSVFYRYAGADNRPAYSKDNKPKPYFAAVSMOGYKADYAMTIGFPGS 276
DB 737 ----NHDDTNRIITYMG---KRP---EGENASYHLAY-----D 765
QY 277 TDYR-----LTSMGVEDRIENENPRI-----EVRGIGQIMKEAMSDAQAT 318
DB 766 KRYTEEREVYSYLRYTGTPIPDNDNKNNSQLVSVAGTVEGTDLSLKFELDLS 825
QY 319 RIKYASKYAQ-----SANWKNSIGMNRGLARLDVIGRRABERAFADIRKN----- 366
DB 826 RPAHGKTEQGLSPKSPKPRATDSGAMSHKLEKADLL--KAIDQOLIAN--VHSDDYFEVI 882
QY 367 --GKSAVYGDVLSLEKAYKEGAKANREMTYLSSETLFGTEVYRFAQFANALATNPDAHA 424
DB 883 DRASDATITDRNGKYVFADKDS--VTLPTQPOVQEFILSGHVRYKE-----KPIQNO 934

QY 464 DIFKNVIDKFKGDTKKYADVFVDSKVYPSDKFHAMIKSM-----D 505
DB 753 DIIDAYCSIGSKDCIKQYKDIYFDE-VMPCKAGEAATFCVYSAPLRANVYCYGOEGG 811
QY 506 KEKFAKA-----LEKDPAYELSKSVYLAARAIADADAMA 538
DB 812 EBAFEKVMGLYLAED--VQLEKGIILFKALACHKVDYTA 846

RESULT 11
US-08-811-583-2
; Sequence 2, Application US/08811583
; Patent No. 6218142
; GENERAL INFORMATION:
; APPLICANT: Massenegger, Michael
; APPLICANT: Riedel, Leonhard
; APPLICANT: Schiebel, Winfried
; APPLICANT: Sanger, Heinz
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
; TITLE OF INVENTION: POLYPEPTIDES HAVING THE ENZYMACTIC ACTIVITY OF AN
; TITLE OF INVENTION: RNA-DIRECTED RNA POLYMERASE (RDRP)
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH & NEAVE
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10020

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/811,583
; FILING DATE: 05-MAR-1997
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley, James F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: MPG-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-596-9000
; TELEFAX: 212-596-9090
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1114 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-811-583-2

Query Match 2.8%; Score 103.5; DB 4; Length 1114;
Best Local Similarity 19.8%; Pred. No. 1.9; Mismatches 229; Indels 207; Gaps 30;
Matches 127; Conservative 78;

QY 92 YGAISQSTVDHDIYLDGVSRTMGE-----ELPIGILSVY-----LRKIYK 134
DB 556 YGIGYGVGVDPDSSGKSLKRSMSKYESDNLIKIDVLGMS-KYQCYLNRQLITLSTIG 614
QY 135 YVDKY-EGOLKITDMEERL---RKAQEVCOELAKKENADENOLCI-----VEPFYS 182
DB 615 YVDEVYLEQOKKBAVDOLDAILHDSLKQAQBALMSGENTILIKMLNCGYKPDAPFLS 674
QY 183 NNEFYLIYDVFK-----DVR---MYFAPSSV-----CK 209
DB 675 -----MMQTRASKLIDLRTSRIFLPGRTMMGCCLDSEFTLEYGOYVQFTAGHGE 728
QY 210 FGGDTIDNMMPRHGTGDFSVFVYAGADNRPAYESKDNKPYKPYFAAVSMOGYKADVDYAM 269
DB 729 FSDDLHPFNNSSTNSNFIKGNVYVAKKPKCLHPGDIRVLKAVNVRALH---HMPD----- 781

QY 270 TIGFP-----GST---DRLYSWGEDRIENENNRIEVRGIKGIMKEAMS- 313
DB 782 CVYFQPKGRPHRPNBCSGSDLDGLDYFCW-----DQMIPRQVQPMETPPASI 832
QY 314 -ADQATRIYKASKYQASANYTKN-SIGMNRGLARLDVIGRKAERAFADMTIRKNG----- 367
DB 833 QLDHDVYTLIEVEBEY--FTNYIVNDSLGI--IANAHVFADEPEPMASDPCKLAELFS 887
QY 368 -----KSAVYGDVLSLE-KAYKEGAKANREMYLSETLFSGTEVVRPAQANLATNP 420
DB 888 IAVDFPKTVPAEITFSQLRPKEYPDEMDKPKDTISYISERVIG----- 929
QY 421 DAHAGILSLDDKYRKYDLPISLDRKYLPAMLDIVRRIIPADKLPIEFKNVIDKFKGDTKK 480
DB 930 -----KLFRKVDKAPQAS-SIATFTROYARSSYDADMEVDFEYIDEAF--DYKT 978
QY 481 YADVFEDKSVYPSYSDKFHML-----KMDKEKFAKALEKRPAYELSKSVYLAARA 531
DB 979 EYDNKLGMLMDYGIKTEKEILSGIMKASKFTDRRKAELAS-----VAARA 1026
QY 532 IOADMANAYALEKGRLEPFAGLREMYPGRALPSDANFTMRASYSIKGYEPQDQAWY- 590
DB 1027 LKKEAR-----AMFRR--NDIDDMLPKASAWYHV 1054

QY 591 -YHTTGKGV---LEKDPKSDEFAVQENILDLFTKKNYGR 626
DB 1055 TYHPTYWGCYNOGLKRAHPISEPMCVYDQLIQIKKDKANR 1095

RESULT 12
US-09-514-302-3
; Sequence 3, Application US/09514302
; Patent No. 6338959
; GENERAL INFORMATION:
; APPLICANT: HATADA, Yuji
; APPLICANT: IGARASHI, Kazuaki
; APPLICANT: OZAKI, Katsuya
; APPLICANT: ARA, Katsutoshi
; APPLICANT: KAWAI, Shuji
; APPLICANT: ITO, Susumu
; TITLE OF INVENTION: GENE FOR ENZYME HAVING BOTH ALKALINE PULLULANASE AND
; TITLE OF INVENTION: ALKALINE ALPHA-AMYLASE ACTIVITIES
; FILE REFERENCE: 2173-105P
; CURRENT APPLICATION NUMBER: US/09/514,302
; CURRENT FILING DATE: 2000-02-28
; EARLIER APPLICATION NUMBER: 08/952,084
; EARLIER FILING DATE: 1997-11-10
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO: 3
; LENGTH: 833
; TYPE: PRT
; ORGANISM: Bacillus sp.
; US-09-514-302-3

Query Match 2.8%; Score 102.5; DB 4; Length 833;
Best Local Similarity 19.0%; Pred. No. 1.4;
Matches 108; Conservative 71; Mismatches 199; Indels 191; Gaps 25;

QY 231 YVAGADNRA-----EYKSDNKPYPYFAAVSMOGYKAD-DYAMTITGPGSITDRILTSMG 285
DB 296 YVAGVKNLPITAIIDEGFNHESGATLEVOARTITSEKADFDWDQSVVYFMLTDRFF----- 353
QY 286 VEDRIENENNRP-----IEVRGIKQ-----GIWKEAMSADQATRI 320
DB 354 ---DGSNNDPHGIGTYDTSKSTYGGGDFKGTITQGLDYLDLDELGITITINISPV-----VDNI 407
QY 321 KYASRYAQSAN-----YWRKNSIG-----MNRGL-ARLDYI--- 349
DB 408 KRDVHSECPDTPYVYAHGYWADNGELNPHFGSMADFHEMIDAHERGIKIMVDVVLNH 467
QY 350 ---GKKRA-----EERA-FADMTIRKNGKSAVYGDVLSLEKAYKBGAKANREMT 394

```

Db 468 TGCGKPCGSSSVANPEPTEDDARFDGMLRDGSGGEVGE-LAGLPDLTENPDVREOV 526
QY 395 Y-----LSETLFGGT-----EVVFAQFANALATNPDAHAGILSLDDKYK 435
Db 527 QWOTDMEKSRKAGNTIDYFRVDYKHYVEDTTWMAFKNALTLTAMPEHKLIGAMGANN 586
QY 436 DYLPISDRKYLPMALDIVRRIRPADKLPDIFKNVIDKKFKGDTKKYADFVDEKSVVPSD 495
Db 587 DDGGLNSGMSDLDLF-----DKNYARDAFQOLDAVOQ 622
QY 496 KFHNA-----MLKSMDEKFAKAEKDPAVELSKSVIAAARLQADAMANAYA 542
Db 623 KLEARNSKLNTATLQOFLSHDEDRFEVEGD---LCKYQVAASLQTLA----- 670
QY 543 IEKGRLEFAGLEBMYPGRALPDSANFTMMSYGSIKGYEPDQGANYNHTTGKGVLEQ 602
Db 671 --KGQPIYTGEEGLPGK---ND-----YPIYTNQN-MPWD 702
QY 603 DPKSDEFAVOENILDFRTKNYGRVAENGOLHIAFLSNNDITGNSGSPYFDKNGRLIGL 662
Db 703 DVDGNELIEHYOKLAFRNDNPTFAKGRKVA---GSDSEGYLLFSRTYGENSVYVGL 759
QY 663 AFDGNWEMSGDIEFEPDLORTISVDIRY 691
Db 760 ----NTEAAKDVTLNFGSSSEAVVTD-RX 783

```

```

RESULT 13
US-09-514-302-2
; Sequence 2, Application US/09514302
; Patent No. 6338959
; GENERAL INFORMATION:
; APPLICANT: HATADA, Yuji
; APPLICANT: IGARASHI, Kazuaki
; APPLICANT: OZAKI, Katsuya
; APPLICANT: ARA, Katsutoshi
; APPLICANT: KAWAI, Shuji
; APPLICANT: ITO, Susumu
; TITLE OF INVENTION: GENE FOR ENZYME HAVING BOTH ALKALINE AND
; TITLE OF INVENTION: ALKALINE ALPHA-AMYLASE ACTIVITIES
; FILE REFERENCE: 2173-105P
; CURRENT APPLICATION NUMBER: US/09/514,302
; CURRENT FILING DATE: 2000-02-28
; EARLIER APPLICATION NUMBER: 08/952,084
; EARLIER FILING DATE: 1997-11-10
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1938
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-09-514-302-2

```

```

Query Match 2.8%; Score 102.5; DB 4; Length 1938;
Best Local Similarity 19.0%; Pred. No. 5.8;
Matches 108; Conservative 71; Mismatches 199; Indels 191; Gaps 25;

```

```

QY 231 VYAGADNRPA-----EYSKNNPKPKPYEAAVSMQGYKAD-DYAMTIGFPGSTDRYLTSMG 285
Db 298 VTGAGVKMLPTTALIDFENRHEGSATLEVOARITTEKADPDMSQVYVFLTRFF----- 353
QY 286 VEDRIENENNR-----IEVGIRK-----GIWKAMASDAQTRI 320
Db 354 --DGDSSNNDPHGIGYTSQSGYGGDEFGITORDLYDELGINTWISPV---VDNI 407
QY 321 KYASKYAQSAN-----YMKNSIG-----NMKGL-ARLDVI--- 349
Db 408 KFDVRSHEGDDPYAYAHGWADNFGELNPHFGSMADFHEMIDAHERGKIKWIDVVLNH 467
QY 350 ---GRKRA-----EERA-FADWIRKNGKSAVGYDVLSELEKAYVEGAKANREMT 394
Db 468 TGCGKPCGSSSVANPEPTEDDARFDGMLRDGSGGEVGE-LAGLPDLTENPDVREOV 526

```

```

QY 395 Y-----LSETLFGGT-----EVVFAQFANALATNPDAHAGILSLDDKYK 435
Db 527 QWOTDMEKSRKAGNTIDYFRVDYKHYVEDTTWMAFKNALTLTAMPEHKLIGAMGANN 586
QY 436 DYLPISDRKYLPMALDIVRRIRPADKLPDIFKNVIDKKFKGDTKKYADFVDEKSVVPSD 495
Db 587 DDGGLNSGMSDLDLF-----DKNYARDAFQOLDAVOQ 622
QY 496 KFHNA-----MLKSMDEKFAKAEKDPAVELSKSVIAAARLQADAMANAYA 542
Db 623 KLEARNSKLNTATLQOFLSHDEDRFEVEGD---LCKYQVAASLQTLA----- 670
QY 543 IEKGRLEFAGLEBMYPGRALPDSANFTMMSYGSIKGYEPDQGANYNHTTGKGVLEQ 602
Db 671 --KGQPIYTGEEGLPGK---ND-----YPIYTNQN-MPWD 702
QY 603 DPKSDEFAVOENILDFRTKNYGRVAENGOLHIAFLSNNDITGNSGSPYFDKNGRLIGL 662
Db 703 DVDGNELIEHYOKLAFRNDNPTFAKGRKVA---GSDSEGYLLFSRTYGENSVYVGL 759
QY 663 AFDGNWEMSGDIEFEPDLORTISVDIRY 691
Db 760 ----NTEAAKDVTLNFGSSSEAVVTD-RX 783

```

```

RESULT 14
US-07-854-596B-40
; Sequence 40, Application US/07854596B
; Patent No. 5434073
; GENERAL INFORMATION:
; APPLICANT: Dawson, Keith M
; APPLICANT: Hunter, Michael G
; APPLICANT: Czaplinski, Lloyd G
; TITLE OF INVENTION: Proteins and nucleic acids
; NUMBER OF SEQUENCES: 73
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dr. John J. McDonnell
; STREET: Ten South Wacker Drive, Suite 3000
; CITY: Chicago
; STATE: IL
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/854,596B
; FILING DATE: 03-JUN-1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: McDonnell, John J
; REGISTRATION NUMBER: 26,949
; REFERENCE/DOCKET NUMBER: 92,337
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-715-1000
; TELEFAX: 312-715-1234
; TELEX: 910-221-5317
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 747 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-854-596B-40

```

```

Query Match 2.7%; Score 101.5; DB 1; Length 747;
Best Local Similarity 19.3%; Pred. No. 1.5; 254; Indels 183; Gaps 33;
Matches 127; Conservative 94; Mismatches 183; Gaps 33;
QY 75 GIVSDGGL-----IFTNHCGYGAIOSQTVHDYLRDGFYSRTWGEPLPPIP---GLS 125

```

```

Db 177 GDTITSOELLAQOSILNKHGTYTYERSSII--VTHNDIFRTI--LPMDOFTYH 230
OY 126 VYLLKRYKVTVDKVGOLKIGIDEMERLKAQVOCGLAKKE-----167
Db 231 VANNQDAVEINK-----SGLNEEINNDLISEKYVLLKGEKPYDPDRSHLKLFTTKY 285
OY 168 -NADENOLCIYEPFYSNNEX---FLIYVDVFNKDVMPVAPSSVKGFG-----GDTDN 216
Db 286 VDVNTNELLKSPQLLTASERNLDPRDLVDPKAKLTY---NNLDAFGIMDYTLGKVED 342
OY 217 WMPPHGTDESFRYACADNPAEYKSKDNKPKYVYPAVSMSQGYKADVTAMTIGFGS 276
Db 343 ---NHDDTNRIITYMG--KRPEGNAS-----YHLAVELQGVPRGSLGVSVAGT 389
OY 277 TDRYLTSMGVEDRIENENNPRIEVRGIGKGIWEKMSADQATRIYASKYAO-----SAN 331
Db 390 -----VEGTNDISLKFELDLTSPRAHGKGTGGLSPKSKP 426
OY 332 YWKSIGNNGLARLDVIGRRRAERAPADWIRKN-----GKSAYYGDVLSLEKA 382
Db 427 FATDSGAPHKLEKADLL--KAIQEQLIAN-VHSNDYFEVIDFASDATTITDRNGKVFYA 483
OY 383 YEGAKKANREMTYLSSETLFGTEVYVRFQAFANALATNPDAHGI---LKSIDDKY-----434
Db 484 DKDG-----SVTL--PTQVQOEFLLSGHVAVRPKPKPIQNAQSKSVYEVTOFT 531
OY 435 -----KDYLPSTL--DRKVLPRAMLDIVRRIRPADKLPDIFKNVIDKKFKGDTKYADFVEDK 488
Db 533 PLNPDDDFRPGJADKDKLTKTL--AIGDITTSOELLAQOSILNKHG-----YTIYENDS 585
OY 489 SVVPPYSDKFHAMLKSMDEKEFAKAIEDKPAVELSKSVIAAARAIQADAMANA-VYIEKNG 547
Db 586 STVTHNDIFRTITLPMDOFTYHVNKRBOAYEINKSGLINEEINNTDLISEKYVLLKGE 645
OY 548 -----RLFPAGLREMYPGALPSDANFT---MMMSYSGIKGYEPODGA--WTN- 590
Db 646 KYDPDFDSHKLFTIKIYDVNTNELKSEQLLTASERNLDFRDL--YDPRKALLLN 703
OY 591 -----YHTTGKGVLEKODPKSDEFAVOENILDLFR--TKNYGR--YAENGOLHIA 636
Db 704 LDAFGIMDYTLGK-----VEDNHDDTNRIITYMGKRPKEGNASVHLA 747

RESULT 15
US-08-417-492-2
; Sequence 2, Application US/08417492
; Patent No. 5750872
; GENERAL INFORMATION:
; APPLICANT: Bennett, Alan B
; APPLICANT: Brummell, David A
; APPLICANT: Grantz, Alexander A
; TITLE OF INVENTION: Nucleic Acids Encoding Ascorbate Free
; TITLE OF INVENTION: Radical Reductase and Their Uses
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew
; STREET: One Market Plaza, Stewart Street Tower
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105-1492
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/417,492
; FILING DATE: 05-APR-1995
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Bastian, Kevin L
; REGISTRATION NUMBER: 34,774

```

```

; REFERENCE/DOCKET NUMBER: 2307E-586US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-543-9600
; TELEFAX: 415-543-5043
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 433 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-417-492-2

Query Match 2.7%; Score 101; DB 1; Length 433;
Best Local Similarity 21.3%; Pred. No. 0.66;
Matches 102; Conservative 62; Mismatches 168; Indels 146; Gaps 23;

OY 220 RYVAGADNRPAEYKSKDNKPKYVYPAVSMSGY-----KADYAMTIGFGSDRRLT 282
Db 22 REFAGQVKGPGELATISKAEVAVYRPPALSKAYLFPPEGAARLPGFHVCGSGG--EROLP 79
OY 283 SMGVEDRIENENNPRIEVRGIGKGIWEKMSADQATRIKYASKYQSANWYKNSIGMNRG 342
Db 80 EMYAE-----KGISLITSTELVAKADLAK--TLVSAGESEFKYQTVIATGTT 125
OY 343 LARLDVIGRRRAERAF-----ADWI-----RKNKSAVYGDVLSLE-----380
Db 126 VKLSDDFGVQADSKNIFYLEIREDADQLVLEALKAKKNGKAVVGGYIGLELSAVLRIN 185
OY 381 -----KAKKEGAKKANREMTYLSSETLFGTEVYVRFQAFANALATNPDAHA- 424
Db 186 NIEVNMVYPEPWCMPLRTTEGIAAFEGYKVK--GVNIK-----GYAVAGDTHPN 236
OY 425 GILKSLDKYKDYLPSTLDRKVLPRAMLDIVRRIRPADKLPDIFKNVIDKKFKGDTKRYADF 484
Db 237 GEVKEVVKL-----DGRVLEA--DLVVGVGARPLTLEKGVVEE--KGIKTDA--283
OY 489 VFDKSVVP---YSDKFHAMLKSMDEKEFAKAIEDKPAVELSKSVIAAARAIQADAMANA 540
Db 284 -FFKTSVPDYAVAGVATFPPLKMYNEIRVEVHDS-----RKSADQAVKAL-----329
OY 541 YAIERKRLFFAGLREMYPGRALPSDANFTMRMSYSGIKGYEPOGAWYNY-HTTGKGVL 599
Db 330 FASEQKGSVDEYDILPYYSRAF-----DLSMGTGDNNGEYVL 368
OY 600 -EKODPKSDEFAVOENILDLFRTKNYGRY-AENGOLHIAFLSNNDITGNGSSGPVDFK 655
Db 369 FGDADPNs-----ATHKFGQYWKDGKIVGAFI-----ESGSPENK 405

```

Search completed: October 17, 2002, 22:41:07
Job time : 32 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 17, 2002, 22:40:26 ; Search time 2436 Seconds
(without alignments)
6116.459 Million cell updates/sec

Title: US-10-008-355-2
Perfect score: 3719
Sequence: 1 MQMKLSILGALLLGASG.....LFMDKMGCCPRLLQELKLI 712

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1797656 seqs, 10463268293 residues
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

Command line parameters:
-MODE=frame+ p2n model -DEV=xlh
-Q/cgn2_1/USPRO_pool/US10008355/runatc_16102002_121153_27693/app_query.fasta_1.903
-DB=genbdl -QWMT=fastap -SUFFIX=trge -MINMATCH=0.1 -DOOPCL=0 -DOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blomsun62 -TRANS=human4.0.cdi -LIST=45
-DOCAALIGN=200 -THR_SCORE=PCT -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=plc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10008355.ecgn1.1.1601_@runatc_16102002_121153_27693 -NCPu=6 -ICPU=3
-NO_XLPHY -NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-TGAPOP=10 -TGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
1: gb.ba:*
2: gb.htg:*
3: gb.in:*
4: gb.om:*
5: gb.ov:*
6: gb.pat:*
7: gb.ph:*
8: gb.pl:*
9: gb.pr:*
10: gb.ro:*
11: gb.sts:*
12: gb.sy:*
13: gb.un:*
14: gb.vl:*
15: em.ba:*
16: em.fun:*
17: em.hum:*
18: em.in:*
19: em.mu:*
20: em.om:*
21: em.or:*
22: em.ov:*
23: em.pat:*
24: em.ph:*
25: em.pl:*
26: em.ro:*
27: em.sts:*
28: em.un:*

29: em.vl:*
30: em.htg_hum:*
31: em.htg_inv:*
32: em.htg_other:*
33: em.htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	907.5	24.4	10689	1 AE004008	AE004008 Xylella f
C 2	186.5	5.0	253217	2 AC016590	AC016590 Homo sapi
3	146.5	3.9	10689	1 AE006549	AE006549 Streptoco
4	144.5	3.9	15569	1 AE000733	AE000733 Aquifex a
5	137.5	3.7	162595	2 OSJN00169	AL662969 Oryza sat
6	133.5	3.6	4004	1 MSP207NAM	X55799 M.sp 207 al
7	133	3.6	10292	1 AE008428	AE008428 Streptoco
C 8	131.5	3.6	11362	2 SPNEU1923	AL449945 Streptoco
9	131.5	3.5	14169	1 AE009858	AE009858 Pyrobacul
10	130	3.5	2652	6 AB1135	AB1135 Sequence 1
11	130	3.5	2652	6 AR065782	AR065782 Sequence
12	130	3.5	2652	6 E26074	E26074 Novel vials.
13	129	3.5	4141	1 CLOCOLA	D13791 C. perfring
C 14	129	3.5	6004	6 BD003707	BD003707 Polynucle
C 15	129	3.5	11280	1 AE007366	AE007366 Streptoco
C 16	129	3.5	26703	3 CBRG45011	AC084652 Caenorhab
C 17	128.5	3.5	291150	1 AP003135	AP003135 Staphyloc
C 18	128.5	3.5	342600	1 AC003363	AC003363 Staphyloc
19	127	3.4	193188	2 AC006884	AC006884 Caenorhab
20	126	3.4	3084	1 LA062096	U62096 Lactobacill
21	124.5	3.3	9389	1 AE008276	AE008276 Agrobacte
C 22	124.5	3.3	11199	1 AE009339	AE009339 Agrobacte
23	124	3.3	36380	3 AF106581	AF106581 Caenorhab
24	123.5	3.3	2142	1 AF039313	AF039313 Moraxella
C 25	123.5	3.3	10296	1 U67538	U67538 Methanococc
C 26	123	3.3	10181	1 AE001761	AE001761 Thermotog
27	123	3.3	340806	1 NMA122491	AL162751 Neisseria
28	122.5	3.3	7009	14 AF2966095	AF2966095 Porcine t
29	122.5	3.3	7858	1 U35629	U35629 Lactococcus
C 30	122.5	3.3	10826	1 SPN49397	U49397 Streptococc
C 31	122.5	3.3	239050	1 AL596169	AL596169 Listeria
C 32	122.5	3.3	299850	1 AP001514	AP001514 Bacillus
33	122	3.3	10698	1 AF218939	AF218939 Bacillus
34	122	3.3	301450	1 AP003185	AP003185 Clostridi
C 35	122	3.3	333050	1 AL596168	AL596168 Listeria
C 36	121	3.3	3225	3 AF438184	AF438184 Toxoplas
37	121	3.3	3516	3 AF438183	AF438183 Toxoplas
38	121	3.3	4419	3 AF006628	AF006628 Toxoplas
39	121	3.3	4535	3 AF006627	AF006627 Toxoplas
C 40	121	3.3	11073	1 AE006589	AE006589 Streptoco
C 41	120	3.2	10380	1 AE000932	AE000932 Methanoba
C 42	118.5	3.2	291050	1 AP000982	AP000982 Sulfolobu
C 43	118.5	3.2	339485	8 AF172282	AF172282 Oryza sat
44	118	3.2	11069	1 AE004124	AE004124 Vibrio ch
45	118	3.2	39937	7 V01146	V01146 Genome of b

ALIGNMENTS

RESULT 1
LOCUS AE004008
DEFINITION Xylella fastidiosa 9a5c, section 154 of 229 of the complete genome.
ACCESSION AE004008
VERSION AE004008.1
KEYWORDS GI:9106961
SOURCE Xylella fastidiosa 9a5c.
ORGANISM Xylella fastidiosa 9a5c

Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group; Xylella.

1 (bases 1 to 10689)

REFERENCE
AUTHORS
Simpson,A.J., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M., Alvarenga,R., Alves,L.M.C., Araya,J.E., Bata,G.S., Baptista,C.S., Barros,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Briones,M.R., Bueno,M.R., Camargo,A.A., Camargo,L.E., Carraro,D.M., Carrer,H., Colauto,N.B., Colombo,C., Costa,F.F., Costa,M.C., Costa-Neto,C.M., Coutinho,L.L., Cristofani,M., Dias-Neto,E., Docena,C., El-Dorry,H., Facincani,A.P., Ferreira,A.J., Ferreira,V.C., Ferro,J.A., Fraga,J.S., Franca,S.C., Franco,M.C., Fromme,M., Furlan,L.R., Garnier,M., Goldman,G.H., Goldman,M.H., Kempers,E.L., Kitzajima,J.P., Ho,P.L., Hohnselt,J.D., Junqueira,M.L., Kempers,E.L., Kitzajima,J.P., Krueger,J.E., Kurama,E.E., Laigret,F., Lopes,M.R., Leite,L.C.C., Lemos,E.G.M., Lemos,M.V.F., Lopes,S.A., Lopes,C.R., Machado,J.A., Machado,M.A., Madeira,A.M.B.N., Madeira,H.M.F., Marino,C.L., Marques,M.V., Martins,E.A.L., Martins,E.M.F., Matsukuma,A.Y., Menck,C.F.M., Miracca,E.C., Miyaki,C.Y., Monteiro-Vitorello,C.B., Moon,D.H., Nagai,M.A., Nascimento,A.L.T.O., Netto,L.E.S., Nhani Jr.,A., Nobrega,F.G., Nunes,L.R., Oliveira,M.A., de Oliveira,M.C., de Oliveira,R.C., Palmieri,D.A., Paris,A., Peixoto,B.R., Pereira,G.A.G., Pereira Jr.,H.A., Pesquero,J.B., Quaglio,R.B., Roberto,P.G., Rodrigues,V., de M. Rosa,A.J., de Rosa Jr.,V.E., de Sa,R.G., Santelli,R.V., Sawasaki,H.E., da Silva,A.C.R., da Silva,F.R., da Silva,A.M., Silva Jr.,W.A., da Silveira,J.F., Silvestri,M.L.Z., Siqueira,W.J., de Souza,A.A., de Souza,A.P., Terenzi,M.F., Truffi,D., Tsai,S.M., Tshabko,M.H., Vallada,H., Van Sluys,M.A., Verjovski-Almeida,S., Vettore,A.L., Zago,M.A., Zatz,M., Zedler,J. and Setubal,J.C.

Direct Submission
Submitted (02-JUN-2000) Organization for Nucleotide Sequencing and Analysis, Bioinformatics Lab - IC/Unicamp, C.P. 6176, Campinas, SP 13083-970, Brazil

TITLE
JOURNAL

FEATURES
SOURCE
1. 10689
Location/Qualifiers
/organism="Xylella fastidiosa 9a5c"
/db_xref="taxon:160492"
/clone="9a5c"
complement(119. 367)
/gene="XF1871"
complement(119. 367)
/gene="XF1871"
/note="hypothetical protein: identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF84677.1"
/db_xref="GI:9106962"
/translation="MTWFLKEDKTEIIDQRIQIAKLISEFAKIQAEIHEVNTTQEM
MKETLKITEESKWPVYVSSIIAAGATATLFIKLPH"
complement(430. 618)
/gene="XF1872"
complement(430. 618)
/gene="XF1872"
/gene="XF1872"

/note="hypothetical protein: identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF84678.1"
/db_xref="GI:9106963"
/translation="MMEBARATAPQGLAASDPQRTIDALQTVSHKRRRAANETASIIQLRQ
IVLRVILVROOKRASP"
941. 1288
/gene="XF1873"
941. 1288
/gene="XF1873"
/note="similar to GI1479340 (percent identity: 36 %/query alignment coverage: 87.8 %/subject alignment coverage: 90.2 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF84679.1"
/db_xref="GI:9106964"
/translation="MSKKITPNIQITGMAGFOFINCKLYTPPERTIEKOLMLSLT
CTLAERWQKMEAEARATAPQGLAASDPQRTIDALQTVSHKRRRAANETASIIQLRQ
VLIRROOKRASP"
1351. 1662
/gene="XF1874"
1351. 1662
/gene="XF1874"
/note="hypothetical protein: identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF84680.1"
/db_xref="GI:9106965"
/translation="MTWFLKEDKTEIIDQRIQIAKLISEFAKIQAEIHEVNTTQEM
NAHINENMARQEMARQKLIETQKSKESOWPVVVASGLAAGATATLFIKLPH"
complement(1839. 2918)
/gene="XF1875"
complement(1839. 2918)
/gene="XF1875"
/note="similar to SP1P03626 (percent identity: 36 %/query alignment coverage: 88.0 %/subject alignment coverage: 105.0 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="phage-related protein"
/protein_id="AAF84681.1"
/db_xref="GI:9106966"
/translation="MLNLVTCGAPGNGKTLVAVDMILROIETDKSLVKSQAVPRSYTD
IEGDEAVNRRLTGYVDSAPEDKRTTQGSYIYDEAHRRFPAGRRERSDDPRCDL
DTHRGGITDLMEVTQWPTKIHHELRLVGEVHLNRAGLQTAGIYKRSADDDPTDI
HOREKAEEVVKFPDRYALVASTLRLVSHKFRPKKWSALSCVIGSVIGLVFMH
YSPDAHLEAASVAGSAGQASLRAAPSSLSVSGMTAVVLETSAPLTSQCV
SSDRSCRCFNTDGYOIDMSVYECRLASPLPFNVYHAYVTSSSAAVSSPSSPS
SSSVSSSVSSVAPLSSSVSSH"
complement(2918. 3320)
/gene="XF1876"
complement(2918. 3320)
/gene="XF1876"
/note="similar to SP1P03625 (percent identity: 36 %/query alignment coverage: 85.1 %/subject alignment coverage: 86.0 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="phage-related protein"
/protein_id="AAF84682.1"
/db_xref="GI:9106967"
/translation="MFSWLAMLRNLNIGQIVARGLVAGLAVTTVPILPIVTSALNL

```

gene
  complement(3195..4412)
  /gene="XF1877"
  complement(3195..4412)
  /gene="XF1877"
  /note="hypothetical protein; identified by sequence
  similarity: putative; ORF located using Glimmer/RBSfinder"
  /codon_start=1
  /transl_table=11
  /product="hypothetical protein"
  /protein_id="AA04683.1"
  /db_xref="GI:9106968"
  /translation="MCSRFFLYFLISALGLTIVISPECLFAONVPVTEVPTSVRY
  SSVISDGVSTAEFARSTALVNGRYVTVVDVDSIGSLAKAVARGMAYNVVS
  MSLGILNGAVGIDELTHEVMGPAIKETIPVCTAMMORRGDGFVSVYPOGLIAR
  INAYOSLSPPQPLVTSSSGSSVSERLHLBEGGERIGHLSKIDSDVPDSSGLP
  SVLITDLAGVRSDDSTVNAVLTDSQGAVALPEIYVSLANKRLSLEDEKASHAP
  DOOPSGGASSPSSSPSSWAFSCMAVVDVDFIDWVSDFLKKPLVPVPPVYDKL
  POAKTWSGLGEGACPPTTPIEFSGKTSVEISYQPCDFALMRPVIVATILA
  AYAGGFRGVKNV"
  complement(4481..4702)
  /gene="XF1878"
  complement(4481..4702)
  /gene="XF1878"
  /note="hypothetical protein; identified by sequence
  similarity: putative; ORF located using Glimmer/RBSfinder"
  /codon_start=1
  /transl_table=11
  /product="hypothetical protein"
  /protein_id="AA04684.1"
  /db_xref="GI:9106969"
  /translation="MKSYIDRGIAFFSLTSSLSIVFADAAAGASPDACAAYALGALS
  GAVALIGAALKAPAIISYGMKMLKAIFG"
  complement(4724..4924)
  /gene="XF1879"
  complement(4724..4924)
  /gene="XF1879"
  complement(4724..4924)
  /gene="XF1879"
  /note="hypothetical protein; identified by sequence
  similarity: putative; ORF located using Glimmer/RBSfinder"
  /codon_start=1
  /transl_table=11
  /product="hypothetical protein"
  /protein_id="AA04685.1"
  /db_xref="GI:9106970"
  /translation="MAOOLMICYQYVVEARVTCQQAAMVPPSLPISEYEDVAILL
  PHIVMCFIVANGFRHFTYVRD"
  complement(4931..5236)
  gene

Alignment Scores:
Pred. No.: 2,47e-54 Length: 10689
Score: 907.50 Matches: 239
Percent Similarity: 49.218 Conservative: 135
Best Local Similarity: 31.458 Mismatches: 291
Query Match: 24.40% Indels: 95
DB: 1 Gaps: 21

US-10-008-355-2 (1-712) x AB004008 (1-10689)
QY 1 MetGIMetLysLeuSerIleLeuLeuGlyAlaAlaLeuLeuGlyAlaSerGly 20
  |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
DB 10499 ATGCGTTTAACTTTCGCTTCCTGTTTACGTAACCTGATCTGCACCTCC--- 10443
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 21 ValAlaLysAlaAspLysGlyMetTrpLeuLeuAsnGlnLeuAsnGlnLysAsnLeuAsp 40
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 10442 ---ACTCATGCGGGCGAAGGCAAGTGCGTCCACAAACACCTT---CCGGAATTGCTGGC 10389
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 41 ArgMetArgGluLeuGlyPheTrpLeuProLeuAspSerLeuTyrSerPheAspLysPro 60
  |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
DB 10388 CCTCTAAAGCAGCGCTGCTACACCTGTCACAGAACATTTGCAATTGACAGCTGAC 10329
  |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
QY 61 SerIleAlaAsnAlaValIlePheGlyGlyCysThrGlyIleThrValSerAsp 80
  |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
DB 10328 CCCATGGGGGCGAGTGTCTCTC-----GGCAATTGACACCGCGAGTCTGGTTTCACCA 10275
  |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|

```

```

QY 81 GlnGlyLeuIlePheThrAsnHisCysGlyTyrGlyAlaIleGlnSerGlnSerThr 100
  |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
DB 10274 GAAGAGCTGGTATCATCAACCATTCAGCTTTCAGGTCGATTCAGTCACTCTACG 10215
  |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
QY 101 ValAspHisAspTyrLeuArgAspGlyPheValSerArgThrMetCylGlnLeuPro 120
  |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
DB 10214 CCTAAAGAACTTAATCAAGAGAGGTTCAATGGCTTCAGTCAGGACGATGACGAGT 10155
  |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
QY 121 Ile---ProGlyLeuSerValIleTyrLeuArgLysIleValIleValThrAspVal 139
  |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
DB 10154 CGGGGTCCCAATCCCGTATTTACGTGCTTGAGCAGATTCAGGATCCGCTCAGAGCT 10095
  |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
QY 140 GlnGlyLeuLeuLysGlyIleThrAspGlnMetGluArgLysAlaGlnGluVal 159
  |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
DB 10094 AAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 10035
  |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
QY 160 Cys-----GlnGlyLeuAlaLysLysGlnAlaAspLysAlaAspGlnLeuGlyIleVal 177
  |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
DB 10034 TTCAGCAACCAAGAAATCCCAATGCGAGAA---GAGCAGGGTATTCGTCGCAATTC 9978
  |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
QY 178 GluProPheTyrSerAsnAsnGluTyrPheLeuIleValIleValAspValPheLysAspVal 197
  |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
DB 9977 TTCAGTTTCGCTGCGAATATCTACAGAGTTCACAGACCTGAGATTCAGAGCAGCTG 9918
  |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
QY 198 ArgMetValPheAlaProPheSerSerValGlyLysPheGlyValAspThrAspAsnTrp 217
  |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
DB 9917 CGTTTGCTTACGACCGACCGAGGAGAGTGGGTAACTTGGGGGCGACGCTCACTAATCTG 9858
  |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
QY 218 MetTrpProArgHisThrGlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsn 237
  |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
DB 9857 ATGTGCGCCGCGATACCGGATCTTCTTCTACCGCGCTACGTCGTTAAAGCGGT 9798
  |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
QY 238 ArgProAlaGluTyrSerLysAspAsnLysProTyrLysProValIlePheAlaVal 257
  |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
DB 9797 AACACGACGAGCTCTCTAAAGAAACATACCTACCGTCCAAACACGCTGAAATTT 9738
  |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
QY 258 SerMetGlnGlyTyrLysAlaAspTyrAlaMetThrIleGlyPheProGlySerThr 277
  |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
DB 9737 AGCGATCAACCATAGTATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 9678
  |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
QY 278 AspArgTyrLeuThrSerTrpGlyValGluAspArgIleGluAsnGluAsnAsnProArg 297
  |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
DB 9677 AATCGTTATGCTTGGTG-----GCCGAATTGAGAAATCTGCTCAG 9636
  |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
QY 298 IleGluValArgGlyIleLysGln-----GlyIleTrpLysGluAla 311
  |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
DB 9635 TGGACTTATCCGGTGTATGCGCAGCATTTCAAGAAATCTGATTCCTTGTAGGCGGCT 9576
  |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
QY 312 MetSerAlaAspGlnAlaThrArgIleLysTyrAlaSerLysTyrAlaGlnIleSerAlaAsn 331
  |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
DB 9575 AGCAACAGAAATCCAGACATTTCAAGTGAATATGCGAGCATTTTACCTGATTCGAATTAC 9516
  |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
QY 332 TyrTrpLysAsnSerIleGlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArg 351
  |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
DB 9515 ACATCGAATAAATCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 9456
  |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
QY 352 LysAlaGlnGlnGluArgAlaPheAlaAspTrpIleArgLysAsnGly----- 367
  |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
DB 9455 AACAACAGTGAAGAAACACACAGTATTACCTGTTGTAACACAGGCGCTACGTGACAC 9396
  |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
QY 368 LysSerAlaValTyrGlyLysPheValLeuSerSerLeuGlnLysAlaTyrLysGlnGlyAla 387
  |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
DB 9395 GAAAGCTTGGCTGCGCATTAACATTTGATTCACGACAAACATATC----- 9348
  |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
QY 388 LysAlaAsnArgGlnMetThrTyrLeuSerGluThrPheGlyGlyThrGlnGluValAl 407
  |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
DB 9347 AAGCAAAATCAGAGATGCTGATTTTGTATGGGTCAAG---TTCACGGAAGATGGTGTACAT 9291
  |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
QY 408 ArgPheAlaGlnPheAlaAsnAlaLeuAla-----ThrAsnProAspAlaHis--- 423
  |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
DB 9290 GGTGTATGCGGTAACTATATGCTTGGCAATCGAGCGACCAAAATATGATGCCCAAGCT 9231
  |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|

```

```

QY 424 -----AlaGlyIleLeuLysSerLeuasp 431
    ||| |||||
Db 9230 GAAGCGGATATCAGAGCGGATCTGCCACCATTTGAAGCACTTAAGCAGATGAG 9171
QY 432 AsplysTyrLysAspTyrLeuProSerLeuaspArglyVal -----445
    ||| |||||
Db 9170 -----CGACGTTACTTACCGAAATGATGCGCCAGATGACGACATCTAGCTGCACC 9120
QY 446 -----LeuProAlaMetLeuaspPileValArgArgGlyIleProAla-----459
    ||| |||||
Db 9119 GAATACACACAGCTTCC-----GTCAACACGCGTGGCGCTGCATTTGAT 9075
QY 460 -----AspLysLeuProaspPilePheLysasnValIleaspLysLysPheLys 475
    ||| |||||
Db 9074 GTTGTGTGGGATGATGATCCCAACCTCTCAAGCGCTTGGGTGACACCAAAATTAGC 9015
QY 476 GlyAspThrLysLysTyrAlaAspPheValPheaspLysSerValValProTyrSerasp 495
    ||| |||||
Db 9014 AGTCTGAGGAAG-----9000
QY 496 LysPheHisAlaMetLeuLysSerMetaspLysGlyLysPheAlaLysAlaIleGlyLys 515
    ||| |||||
Db 8999 -----CTCAAGTGTATTATCCCGACCGCTGCACCTTGAAGACAGTCA 8955
QY 516 AspProAlaValGlyLeuSerLysSerValIleAlaAlaArgAlaIleGlyAlaasp 535
    ||| |||||
Db 8954 GATCCAGGATCCGTTATGCTGTGCGCATTCATGCGACATTCAGATTTAGAGCTGAG 8895
QY 536 AlaMetAlaAsnAlaTyrAlaIleGlyLysGlyLysArgLysPhePheAlaGlyLeuArg 555
    ||| |||||
Db 8894 AATAAATCCGTACTGAGAACTCTTAACGCGGCTCGATCTACTTGCAAGCTTACT 8835
QY 556 GluMetTyrPro-----GlyArgAlaLeuProSeraspAlaAsnPheThrMetArgmet 573
    ||| |||||
Db 8834 GACTACACACAGAGCATGAGCAAAATTGTTTACCAGAGCGCACTCTATTACGATTT 8775
QY 574 SerTyrGlySerIleLysGlyTyrGluProGlnaspGlyAlaTyrTyrAsnTyrHisThr 593
    ||| |||||
Db 8774 ACCTTCGCGCCAGCGTGAAGGATATTCACCAAGATGCGTGCAGTACACACCGCTTTCG 8715
QY 594 ThrGlyLysGlyValIleGlyLysGlnaspProLysSeraspGluPheAlaValGlnGlu 613
    ||| |||||
Db 8714 ACACGCGAGGCTGTCATGCGCAAG---AATACAGCGCTGAGCCATTGATTCACCGAAG 8658
QY 614 AsnIleLeuaspLeuPheArgThrLysAsnTyrGlyArgAlaGluAsn-----Gly 631
    ||| |||||
Db 8657 TCACGATTAATGCGATTAAGCAAGACAGATCCCAATTAGCCGATCAGCGTATTGCG 8598
QY 632 GlnLeuHisIleAlaPheLeuSerAsnAspIleThrGlyLysSerGlySerPro 651
    ||| |||||
Db 8597 ACTGTCCTGTGAATTTCTCTGATCTGACATTTCTGCGCGCACTCGGATCGCA 8538
QY 652 ValPheaspLysasnGlyArgLeuIleGlyLeuAlaPheAspGlyAsnTyrGluAlaMet 671
    ||| |||||
Db 8537 GTCATTAGTTCGCCACGCAAAATTTGGTGGTTTACCATTTGACGGTAATTGGAGCTCTC 8478
QY 672 SerGlyaspIleGluPheGluProaspLeuGlnArgThrIleSerValaspIleArgTyr 691
    ||| |||||
Db 8477 AGTAGTAATGCGTGTGATCTGATGATGACGCAACATTCAGATTGATGCGTAT 8418
QY 692 ValLeuPheMetIleaspLysTyrGlyLysProAlaGlyLeuIleGlnGlyLeuLysLeu 711
    ||| |||||
Db 8417 GTGCAGTGATCATGACTGAAAGTTGACACACACCATCTACTAAAGAAATTGAATTGG 8358

RESULT 2
AC016590/c 253217 bp DNA linear HTG 20-APR-2001
LOCUS Homo sapiens chromosome 19 clone CTD-3220F14, WORKING DRAFT
DEFINITION AC016590
SEQUENCE, 39 unordered pieces.
ACCESSION AC016590
VERSION AC016590.6 GI:13699590
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE human.

```

```

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 253217)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 19
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 253217)
AUTHORS DOE Joint Genome Institute.
JOURNAL Direct Submission
COMMENT Submitted (04-DEC-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Apr 20, 2001 this sequence version replaced gi:7711568.
-----Genome Center
Center: Joint Genome Institute
Center code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 940643, BC905667
Center clone name: CIRB-EL_3220F14
-----
Summary Statistics
Consensus quality: 207648 bases at least Q40
Consensus quality: 225569 bases at least Q30
Consensus quality: 232428 bases at least Q20
Estimated insert size: 250510; agarose-ef estimation
Estimated insert size: 249417; sum-of-ctgls estimation
Quality coverage: 9.03 in Q20 bases; agarose-ef estimation
Quality coverage: 9.07 in Q20 bases; sum-of-ctgls estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 39 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1019: contig of 1019 bp in length
1020 1119: gap of unknown length
1120 2152: contig of 1033 bp in length
2153 2252: gap of unknown length
2253 3701: contig of 1449 bp in length
3702 3801: gap of unknown length
3802 5053: contig of 1252 bp in length
5054 5153: gap of unknown length
5154 6184: contig of 1031 bp in length
6185 6284: gap of unknown length
6285 7809: contig of 1525 bp in length
7810 7909: gap of unknown length
7910 8986: contig of 1077 bp in length
8987 9086: gap of unknown length
9087 10090: contig of 1004 bp in length
10091 10190: gap of unknown length
10191 11225: contig of 1035 bp in length
11226 11325: gap of unknown length
11326 12365: contig of 1040 bp in length
12366 12465: gap of unknown length
12466 14535: contig of 2070 bp in length
14536 14635: gap of unknown length
14636 16488: contig of 1854 bp in length
16489 16589: gap of unknown length
16590 19109: contig of 2520 bp in length
19110 19209: gap of unknown length
19210 21841: contig of 2632 bp in length
21842 21942: gap of unknown length
21943 23735: contig of 1794 bp in length
23736 23835: gap of unknown length
23836 25961: contig of 2026 bp in length
25962 29117: contig of 3156 bp in length
29118 29217: gap of unknown length
29219 31907: contig of 2690 bp in length

```

```

* 31908 32007: gap of unknown length
* 32008 36364: contig of 4357 bp in length
* 36365 36464: gap of unknown length
* 36465 43309: contig of 6845 bp in length
* 43310 43409: gap of unknown length
* 43410 48185: contig of 4776 bp in length
* 48186 48285: gap of unknown length
* 48286 55626: contig of 7341 bp in length
* 55627 55726: gap of unknown length
* 55727 65781: contig of 10055 bp in length
* 65782 65881: gap of unknown length
* 65882 69673: contig of 3792 bp in length
* 69674 69773: gap of unknown length
* 69774 79389: contig of 9616 bp in length
* 79390 79489: gap of unknown length
* 79490 91453: contig of 11964 bp in length
* 91454 91553: gap of unknown length
* 91554 96671: contig of 5118 bp in length
* 96672 96771: gap of unknown length
* 96772 100544: contig of 3773 bp in length
* 100545 100644: gap of unknown length
* 100645 111423: contig of 10779 bp in length
* 111424 111523: gap of unknown length
* 111524 115278: contig of 3755 bp in length
* 115279 115378: gap of unknown length
* 115379 124011: contig of 8633 bp in length
* 124012 124111: gap of unknown length
* 124112 138475: contig of 14364 bp in length
* 138476 138575: gap of unknown length
* 138576 140613: contig of 2038 bp in length
* 140614 140713: gap of unknown length
* 140714 144949: contig of 4236 bp in length
* 144950 145050: gap of unknown length
* 145051 164352: contig of 19303 bp in length
* 164353 164452: gap of unknown length
* 164453 170540: contig of 6088 bp in length
* 170541 170640: gap of unknown length
* 170641 185046: contig of 14406 bp in length
* 185047 185146: gap of unknown length
* 185147 222209: contig of 37063 bp in length
* 222210 222309: gap of unknown length
* 222310 253217: contig of 30908 bp in length.
* 222310 Location/Qualifiers

```

```

FEATURES
source 1..253217
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="19"
        /clone="CFD-3220F14"
        /clone_id="Caltech human BAC library D"

```

```

BASE COUNT 62246 a 62867 c 62532 g 61752 t 3820 others
ORIGIN

```

Alignment Scores:

```

Pred. No.: 0.1 Length: 253217
Score: 186.50 Matches: 70
Percent Similarity: 47.98% Conservative: 49
Best Local Similarity: 28.23% Mismatches: 107
Query Match: 5.01% Indels: 22
                Gaps: 6

```

```

US-10-008-355-2 (1-712) x AC016590 (1-253217)

```

```

Oy 483 AsphPheValPheAspSerValProTyrSerAspPheHisAlaMetLeuLys 502
    ||| : : : : : ||| : : : : : |||
Db 4997 GATGGCAGTCGCAAGCGGACCTGGTTCGCCCTGCAACAACTGTATGCGGCGCAAG 4938
    ||| : : : : : ||| : : : : : |||
Oy 503 SerMetAspLysGluTyrPheAlaLysAlaIleGluTyrAspProAla---ValGluLeu 521
    ||| : : : : : ||| : : : : : |||
Db 4937 CTGGCGGCAAGGCGGCGGACCTGGCTGTAACGACGCGCGCGGAATTCAGGCGC 4878
    ||| : : : : : ||| : : : : : |||
Oy 522 SerLys---SerValIleAla-AlaAlaArgAlaIleGluAlaAspAlaMetAlaAsnAl 540
    ||| : : : : : ||| : : : : : |||
Db 4877 AGCAAGACAGCTTCATCTAGCGCGCGCGCTGATGTACCGCGTATTCGCTGAAC-- 4820
    ||| : : : : : ||| : : : : : |||

```

```

Oy 540 atyAlaIleGluLysGlyLysArgLeuPheAlaGlyLeuArgGluMetTyrPro-- 559
    : : : : : ||| : : : : : |||
Db 4819 -AAGCTAATGGGAAAGACTGAAGACTGGCGGCTCAGATCCAGCAAGATGACCAAC 4761
    : : : : : ||| : : : : : |||
Oy 560 -----GlyArgAlaLeuProSerAspAl 567
    : : : : : ||| : : : : : |||
Db 4760 TACATGAAGCAAGATGCGCTTCACCTGAGAGCAAGAGGCGCAACCGCTCTCTGACGC 4701
    : : : : : ||| : : : : : |||
Oy 567 aAsnPheThrMetArgMetSerTyrGlySerIleLeuLysGlyTyrGluPro---GlnAspG1 586
    ||| : : : : : ||| : : : : : |||
Db 4700 CAACAGCACCTGGCGCGGACCTTCGCGCGGCGGCGCGCGCGCGCGCGCGCGCGCGG 4661
    ||| : : : : : ||| : : : : : |||
Oy 586 yAlaTyrPyrAsnTyrHisThrThrGlyLysGlyValLeuGluLysGlnAspProLys-S 606
    ||| : : : : : ||| : : : : : |||
Db 4640 CACCACCTGACCGCGCTGACACCGCTAATGCGCTGTATGACCAAGCGCGCGCGGAG 4581
    : : : : : ||| : : : : : |||
Oy 606 erAspGluPheAlaValGlnGluAsnIleLeuAspLeuPheArgThrLysAsnTyrGlyA 626
    : : : : : ||| : : : : : |||
Db 4580 GCGAGTTCATGACGACGACGAAACAACTGCGCGCTGATCATGCGCTGACCTCGCA 4521
    : : : : : ||| : : : : : |||
Oy 626 rGlyTyrAlaGlnAsn-----GlyGlnLeuHisIleAlaPheLeuSerAsnAspIleT 644
    : : : : : ||| : : : : : |||
Db 4520 AATTTGTCGATCCGAAAGCTGAAACCGTGGCGGCTGACTACCTGCGCACCTGACATCG 4461
    : : : : : ||| : : : : : |||
Oy 644 hrGlyLysAsnSerGlySerProValPheAspLysAsnGlyArgLeuIleGlyLeuAlaP 664
    ||| : : : : : ||| : : : : : |||
Db 4460 CCGGCGGCACTCCGCTTCGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCT 4401
    ||| : : : : : ||| : : : : : |||
Oy 664 heAspGlyAsnTyrPheAlaMetSerGlyAspIleGluPheGluPro-AspLeuGlnArg 683
    ||| : : : : : ||| : : : : : |||
Db 4400 TCGAGCGGACCGCTGACTGATTCATTCCTGCGGACTGACCAAGCGCGCAACCGCT 4341
    : : : : : ||| : : : : : |||
Oy 684 ThrIleSerValAspIleArgTyrValLeuPheMetIleAspLysTyrGlyGlnCysPro 703
    : : : : : ||| : : : : : |||
Db 4340 TCGATCCAGGTCGACCTGCGCTGATCTGTGGAACATGAACAGCTGACCGACGAC 4281
    : : : : : ||| : : : : : |||
Oy 704 ArgLeuIleGlnGluLeu 709
    ||| : : : : : ||| : : : : : |||
Db 4280 AACCTGCTGAAGAATG 4263
    : : : : : ||| : : : : : |||

```

```

RESULT 3
AE006549 10889 bp DNA linear BCT 01-JUN-2001
LOCUS
DEFINITION Streptococcus pyogenes M1 GAS strain SF370, section 78 of 167 of
the complete genome.
ACCESSION AE006549 AE004092
VERSION
KEYWORDS
SOURCE
ORGANISM Streptococcus pyogenes M1 GAS.
Streptococcus pyogenes M1 GAS.
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
Streptococcus.

```

```

REFERENCE
AUTHORS 1 (bases 1 to 10889)
Ferretti,J.J., McShan,W.M., Adjie,D., Savic,D., Savic,G., Lyon,K.,
Primeaux,C., Sezate,S.S., Surorov,A.N., Kenton,S., Lai,H., Lin,S.,
Qian,Y., Jia,H.G., Najjar,F.Z., Ren,Q., Zhu,H., Song,L., White,J.,
Yuan,X., Clifton,S.W., Roe,B.A. and McLaughlin,R.E.
Complete genome sequence of an M1 strain of Streptococcus pyogenes
Proc. Natl. Acad. Sci. U.S.A. 98 (8), 4658-4663 (2001)

```

```

TITLE
JOURNAL MEDLINE
PUBMED 21192684
11296296
2 (bases 1 to 10889)

```

```

REFERENCE
AUTHORS Ferretti,J.J., McShan,W.M., Adjie,D., Savic,D., Savic,G., Lyon,K.,
Primeaux,C., Sezate,S.S., Surorov,A.N., Kenton,S., Lai,H., Lin,S.,
Qian,Y., Jia,H.G., Najjar,F.Z., Ren,Q., Zhu,H., Song,L., White,J.,
Yuan,X., Clifton,S.W., Roe,B.A. and McLaughlin,R.E.
Direct Submission
Submitted (10-APR-2001) Department of Microbiology and Immunology,
University of Oklahoma Health Sciences Center, 940 St Young Blvd,
Oklahoma City, OK 73104, USA

```

```

FEATURES
source 1..10889
Location/Qualifiers

```

```
gene
/organism="Streptococcus pyogenes M1 GAS"
/strain="SF370"
/serotype="M1"
/db_xref="taxon:160490"
CDS
98..739
/gene="Spy1039"
98..739
/gene="Spy1039"
/note="Best Blastp hit = no blast hits"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK33931.1"
/db_xref="GI:13622188"
/translation="MTKQIEKLADLPPEAATLYLDMIPEREGYASYVYOTPKODI
OKRRIKQLTKNEPRLDTPSYVMPTKTAEQDLADKGLAYLLSTVANQYFAAY
NEDOMFOYLKNGARAVAFYLDKGEFLEMTGILETYENENPRAIVAVQYLYLAD
MOKRLPSLSIMVQESSKLFYDQGFVLAKGCKPENNDSDVCLVKYFSKP"
802..1932
/gene="hemN"
/note="Spy1040"
802..1932
/gene="hemN"
/BC_number="1.3.3.3"
/note="Best Blastp hit = pir||B69640 coproporphyrinogen
oxidase (EC 1.3.3.3) III, oxygen-independent hemN -
Bacillus subtilis >g11303805|dbj|BAA12461.1| (D84432)
Yger [Bacillus subtilis] >g11263496|emb|CAB14492.1|
(299117) coproporphyrinogen III oxidase [Bacillus
subtilis]"
/codon_start=1
/transl_table=11
/product="putative coproporphyrinogen III oxidase"
/protein_id="AAK33932.1"
/db_xref="GI:13622189"
/translation="MSKKPTSAVVIHPCTQICVYCDPSKVFIONQPNVAYLKALQF
FDSYGRDLAKTLYIGGTPATITAKOLEVILNLENLDDLEETITANFEDLPE
KIAVLORSVAVNRISLIGVOTFNNKQLQIGRSHDEQIYSTIANLKTAGFHNISDILY
ALPGQLDQKENVAKALADLIPHLQSLILEHTVFNKMRGLNPTDLEEM
FEVILISEMANGENHEIENFTKPGFESRHNLMYMDNVEYFCGAGASGYLNGIRON
RVPDIOHLYKAVEANGARLNDEVLKREEMEEELFLGRKKTGVSIQRFEGKGMSPFE
RGNIVRELONGLVAKDDAFVMTKKGLFGDSVAERFLD"
2126..2695
/gene="Spy1042"
2126..2695
/gene="Spy1042"
/note="Best Blastp hit = gb|AA071730.1| (U65643) acyl-ACP
thioesterase [Myristica fragrans]"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK33933.1"
/db_xref="GI:13622190"
/translation="MRSFLACHIFKPSPLKOKPEPIINFVIANFIMIRGGILY
DILAYFALLNPOTRKAYATIPEDIVAFEPEDFYKALKLRVPMPELSDSDRDYVRFD
IDMNGHVNNSKTLDMYDVLGCEFLKTHQPLKMTLKRVESVPGQITSDYHLDQLS
YQIQTSDGQLNQAMLEMRARIKQTESETD"
2695..3459
/gene="Spy1043"
2695..3459
/gene="Spy1043"
/note="Best Blastp hit = pir||H70023 N-acetyl-glucosamine
catabolism homolog ynfF - Bacillus subtilis
>g112633726|emb|CAB15219.1| (Z99120) similar to
N-acetyl-glucosamine catabolism [Bacillus subtilis]"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK33934.1"
/db_xref="GI:13622191"
/translation="MPYKGYLDLQGTIYQGNRIRPAGRFRIKRLQERGIPLLYTNN
TRTPBMVOSMLANOPHVTESIETITATMAVADVINDNRKTAIVAGIETLKSATA
AAGYVEELENPAYVVGIDSOVYTEMALATLALIGALFIGNPDLNIPTEGRIMP
```

```
gene
AGALLNLEAATRVKPFVFGKPNATIMNKSLEVLGIORSEAVVWGDNVLTIDMAGION
DIATILVTGTRPEEVPTLPIDPHVLVSLDEMR"
3459..4091
/gene="Spy1044"
3459..4091
/gene="Spy1044"
/note="Best Blastp hit = sp|P54589|YHCE_BACSU HYPOTHETICAL
29.5 KD PROTEIN IN GLPD-CSPB INTERGENIC REGION
>g117475050|pir||A69822 hypothetical protein yhcE -
Bacillus subtilis >g11239980|emb|CAB5668.1| (X86983)
hypothetical protein [Bacillus subtilis]
>g112633228|emb|CAB12733.1| (Z99108) yncE [Bacillus
subtilis]"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK33935.1"
/db_xref="GI:13622192"
/translation="MVENKLLCSWMLALITITVSTWLMYPLEVDHLKEQVVF
MSKDAILHNYNGLNLYLNPVYRLFEAFHSSADGLKHFADVKMLFLTQVYFLGLL
YPLKTFQRLTKRPMLOKPLILALFLEPILMIGMASPIGFEPHPTLEHOVLFGDS
SWLFDPKDSYVIMLPEVFLHCFLEFMIVYEIILMSLVGLARMQRLANE"
4571..8677
/gene="Spy1046"
4571..8677
/gene="Spy1046"
/note="Best Blastp hit = pir||H81982 hypothetical protein
NMA0631 [imported] - Neisseria meningitidis (group A
strain 22491) >g117379363|emb|CAB83920.1| (A162753)
hypothetical protein NMA0631 [Neisseria meningitidis]"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK33936.1"
/db_xref="GI:13622193"
/translation="MDKRTSISLDITGTSVGVAVITDEKYSKFKVYLGNDRSIK
KNLIGALLDSEGETEATRLKRTARRRTYRRNRICVLOEITSNEAKVDOSFRLKE
ESFLVEEDKRHRHPFNGNIVDEVAIHEKYPPIYHLRKLLVSTQKADRLYLALAH
MTRFNRHPIEDLDNDSDVDKFLQIOLQVOTNOLEENPINAQVDAAILIARLSK
SRLEENLILQGEKKNGFLGMLIALSLDLTFNFSNPFLLADAKLOLSKOYDODLD
NLLAOLGDDYALFLAAKNLSDAIILSDILRVNTEITKAPLSASMKRYDEHODLT
LKAIYRQOLPERKYELEFPDQSKNGAGYIDGASQDEPEFKPILEKDCREELLYK
LNRDLKQRTFDNGSTPHQIHGELALILRQEDFPFLKDNREKELITFRPIY
YVGPLARGNRFRAMTRKSEETITPWNVEEVAVDGKASQFIERMTNEDKLNPKSVL
PKHSLYEYFYVNELTKVYVTEGRKPAFLSGDKKAIYDLKRTNKKYVQOLKE
DYFKRTCEQDSVEISGVEDRFNASLGTYDILKIKDKQFLDNEMEDLIEDIYVLT
LFEDRPMTEERLKTVAHLFDQVMKOKRRRTGSGRSKRLINGIRKQSGSTLIDF
LKSDFPAMNFMOLIHDSLTFRKEDIOKROVSGCDSJAEHIANAGSPAIKKGILQF
YAVUDELYKVGRRHKRPENIYEMARENQTOGQKNSRBRMRTEGRIKELQSILKE
HPVENTQLONEKLYLYLQNGRDMYVDOELINRLSDVDVHIVQSFLEKDSIDNKV
LIRSDKNRGSNDVNSEEVYKMKKNYMRQLNNAKLITOKRFPNLKAEKRGSEIDKA
GFIKROLVETROITKHAVAOILDSRMKTYDENDKILREKVVTLKSKLVSDRKEQF
YKVEILNYYHHAHDVLYNNAVGTALIKKPKLESEEVVDYVYVVRKMYIASBOEIG
KATATYFYSNINMNFKITITLANGEIRKRLPIETNNGEGLVPMOKGRFAYRVRLS
MPQAVNKRTVEQVTOGFSKESLTPKRNKSLIARKKDPKATVGGDFSTVAYSLV
AVEVNGSKKRLKSVKELGILTIMERSSEKPNIDLEAAGYKVRKDLILKPKSLF
ELENGKRMLASAGELOKGNELALPSKYVNFYKLSHEKLEKSEPNOKOLFVEOH
KHVDEITIOISEFSKRYLADANLDKYNVAFNKLRSRDKIRFOAENIHLFLTNLGA
PAAFKVFDTTIDRKRTYSTKEVLDNLTILHQSITGLYERIDLSQLGCD"
8677..9546
/gene="Spy1047"
8677..9546
/gene="Spy1047"
/note="Best Blastp hit = pir||B81299 hypothetical protein
Cj1522c [imported] - Campylobacter jejuni (strain NCTC
11168) >g116968946|emb|CAB73942.1| (A1139078) hypothetical
protein Cj1522c [Campylobacter jejuni]"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK33937.1"
/db_xref="GI:13622194"
/translation="MAGWRTVVVNTHSKLSYKNNHLIFKDAYKTELILHLSIEDILLLE
```


Oy	441	uasprgysvalleupro-----Alamtleusplilevalirgaravag-----	456
Db	7816	AGTCGCGAAAGTATTGTGCCATCCGCCCAAGTGCAATTTTCTCAGAAAACAGAAGTACACAC	7875
Oy	457	-----llepro-----Alasplylsleuproaspil	465
Db	7876	AGGCAGATTCTCCAAAGAGTAGCAATTTTACCMAAAGAAATTCGGACAAGCTT-----	7927
Oy	465	ePhelysnnvallileasplylslyspheylsgilyasphrhtyrsystyrAliaaspheya	485
Db	7928	-----ATTGCTGCTTAATAAAAAAGACTGGGATCCAAAAAATATGTGTGT-----	7969
Oy	485	lPheasplysserValvalProtyrSerAsplysSphehisla-----	499
Db	7970	-TTTATAGTGTCCAACGCGTAGCTTATTCAGTCTCTAGTGTGTCTGAAGTCGCAAAAAAGGAA	8028
Oy	500	-----MetleuysSsr-----MetaAsplysGilly	508
Db	8029	ATCCGAAGAAGTTAAATATCCGTTAAAGAGTAGTACTAGGAGTACCAATTTATGAAAGACAGTTC	8088
Oy	508	sPheha-----lysAlalIegilylas	516
Db	8089	CTTTGAAAAAATCCCATGTGACTTTTGAAGACTTAAGCATATATAGGAAGTTAAAAAAGA	8148
Oy	516	pProcalavalglulseSerlysservalilaialaialargalatIegilAliaaspai	536
Db	8149	CTTATATCTTAAACACTTAA-----	8170
Oy	536	ameAlaasnAlaryAlaIlegilylgilylsyargLeuphehalagilyleuaragl	556
Db	8171	----TATAGTCTTTTTGGAGTTAGAAAAGCGTCGTAAACGAGTGCCTAGTCCGCGAGA	8226
Oy	556	umetyrpfroglyarg-----Alaleuproserasp---AlasnpethmetArgme	573
Db	8227	ATTACAAAAGAAATGAGCTGGCTGTCCCAAGCAAAATATGTGAATTTTATATTTAGC	8286
Oy	573	tSer---TyrglyserilelysgilytyrGluproInaspely-----Alaty	588
Db	8287	TAGTCATTATGAAAAGTTGAAGGGr---AGTCCAGAAGATTAACGACAAAACAATTTGT	8343
Oy	588	pTyrsAntYrhIsthrtHrNgilyLysgilyValleugilyLysglinspPolysSeraspCi	608
Db	8344	TGTGAGCAGCAGAT-----AAGCATATTATTAGATGAGATTATGAGCAAATCAGTGA	8394
Oy	608	uphealavalaln-----	612
Db	8395	ATTTCTTAAGCGTGTTATTTTAGCAGATGCCAATTTAGATAAAGTCTTACTGATATAA	8454
Oy	613	-----glusnilleuaspLeupeargTh	621
Db	8455	CAACATAGACACAAACAAATACGCGAACACAGCAAAATATTATTCATTATTTACTGT	8514
Oy	621	rlyshsnTyrgly-----Ar	626
Db	8515	GACGACATCTTGAGAGCTCCCGCTGCTTTAATATTATTGATCAACAATTTGATCGTAACG	8574
Oy	626	qTyraIaeIusngilylnleuHisIIealarpheuensermsnaspiIetHgly	645
Db	8575	ATATACGCTTAACAAGAAGTTTAGATGATCCACTTATTCATCATTCATCACACTGGr	8632
RESULT 4	AE000733	1556 bp DNA linear BCT 25-MAR-1998	
LOCUS	AQU00733	Aquifex aeolicus section 65 of 109 of the complete genome.	
DEFINITION	AQU00733 AE000657		
ACCESSION	AE000733.1 GI:2983720		
VERSION	AE000733.1		
KEYWORDS			
SOURCE	Aquifex aeolicus.		
ORGANISM	Bacteria; Aquificales; Aquificaceae; Aquifex.		
REFERENCE	1 (bases 1 to 15369)		
AUTHORS	Decker,L.G., Warren,P.V., Gaasterland,T., Young,W.G., Lenox,A.L.,		

[illegible]


```

gene
/4739..5209
/gene="aq_1295"
/4739..5209
/gene="aq_1295"
/codon_start=1
/transl_table=11
/product="putative protein"
/protein_id="AAC07293.1"
/db_xref="GI:2983731"
/translation="MKKEETRTGIGLKDVSIIRLVSPILLPVGGFVCEYGIPIRYIK
IDENEYKLEISYVEKLNLSVNCNIEQVYDVKYEELRYALPEGVFTIARYANGITLR
ELSYLEIYTSVGKGIIDEDRELDOEPAPVKNRETIIGTIREVILLLEIMYS"
5306..7738
/gene="clpc"
/note="aq_1296"
5306..7738
/gene="clpc"
/codon_start=1
/transl_table=11
/product="ATP-dependent Clp protease"
/protein_id="AAC07290.1"
/db_xref="GI:2983728"
/translation="MPEKTEPAROVILRARERALELGHSLGSEHLLALIDEDIP
VLVSLKLVSPKIRKAIMGOITKSHSGSVLPAPDKRRIEFAVEARLHLHQFQIP
EHLILVNERKTLGIRILNSFLDIEYVAREVLQQLLGLPEQSQRVSTPMEDRS
RDLTOMAREGLDIPVIGREKEIEIRVIOIIVRRKNMVLGLDGPVKTAVELAOHI
ANKEVEPILNKRVALDLAALVAGTIRGOEFERIKNLIKLEKANNVLLFIDELHT
LVNGSAEGSIDASNMKLPAALARGEIOVIGATVDEYRKYTEKDGALEREFOVYEO
PSPEDTIOLILYGLKPKPEEFHKVETPEALEKAVELSVKVIIDPNRDKIDMDAG
AUKIKATQIPPELKEIEBKIKLEKEKEKAYVEQDYERAAKIRDEBLIRAKFETIK
AKKKEHEKRPKRVDEDAVAVARMTGVFVKRVHSDAKLLHIEBELKRVYGDDE
ALAAVAKAIRSRVGLGKRRPIGVEMFLGPTGVGRTEAKALAEYLFQTEADLIFRD
MSEYMEKHTVSRIGAPPGVYEGGOLTEAVRRRPYSVLDEIEKAPDVNFIL
QIPDDGLITDGLRTVFSNTIIMTSLNGARLITGTMGFEKKOGLNFEERKNVL
DOVKTFSPEFLNRDEIIVFRLEKKDIEQILLOLKEINERKIDMNVVRKLHKSIV
NMIIIEKYKPEFGARSLKRALQNRHVEDLAEELIKGTAKGVEEVEIRVRNDRKYLAPV
KKKEKVALIEK"
7776..10112
/gene="omp"
/note="aq_1300"
7776..10112
/gene="omp"
/codon_start=1
/transl_table=11
/product="outer membrane protein"
/protein_id="AAC07292.1"
/db_xref="GI:2983730"
/translation="MKKEPILALFLGVANAARIEKTIKGNKYIPDELIRGLLSVRK
GSEFSLERIKDILRKILFRGILTKVEAYEEDONKNGPELVIEDVPVYIKIEFBN
EEFSDDLKELYLIELELGEVEEETIEGTSPPALEERLAVKQOLKGLIIFQIE
SLRRRIIEVYKRGHFGVTKVEYKILPKKSGSLPFIHBEKLYTSDIEYGNKHS
KUKGIMELKEPNIPLFRLHPAFSEETLKEDVKIEEYNEEFLYSDIEYGEKHS
ARKVIYKVGSPRYLKEIKIEGNTLFAVSELDNLIKKEKRGRRFRRIYKINR
IRREYAEIGFLNLSVEREIVNPEKKEVSELTITIGKPYVKKIKIKYEREDVYI
RRMRQENELAKKGIENSKTRIMNIGYEDVQIEPVRDAMWDLVAKIRERFTQ
FSVGLSYNEVYGLAGFTELKGNFREGTDLAGISVSGLSYRNNAISYRWKRLKRV
DLDLASFDRILEYDYVERTGFSVALSKELSEYWRASIGTSIORVYSDIDEASY
VKBOAGRDRSKIFETLITRDYLLPTGSLFVRNSVGLLGGDEKFFYFEVGG
AKYSEDPYFDGTIILSLKGEVFEVGGKKVPIIDREFVGGDSIRGYGYAGAPVD
PMTLDLIGAKKVIITSELMVYVFKMLYFAGFPDYGIDGADKSDRKLSTFRGYSIG
VRITTFPAPRIRIMARKTKVPEPDITNRSRIHFVLGTFP"
complement(10138..10350)
/gene="cspc"
/note="aq_1303a"
complement(10138..10350)
/gene="cspc"
/codon_start=1
/transl_table=11
/product="cold shock protein"

```

```

gene
/protein_id="AAC07291.1"
/db_xref="GI:2983729"
/translation="MSFRGYKRFKFSODKGGITREDTADNVFHTDIOMESGKTLQ
KGGKVEFDVEDTKGPPAKNVRVLGE"
complement(10454..11170)
/gene="hisa"
/note="aq_1303"
complement(10454..11170)
/gene="hisa"
/codon_start=1
/transl_table=11
/product="phosphoribosylformimino-5-aminolmidoazole
carboxamide ribotide isomerase"
/protein_id="AAC07284.1"
/db_xref="GI:2983722"
/translation="MNLKEITIPALIDMEGKAVRLYKGFNKVYSERPWEIKNFS
DLGFKRLHVVDLGEAGGRKPNLEVRIRIARENFEGEVGGIRSYEVALDEGID
FVIGITLAVKNKEEFLKILENPNRVLIAIDSKOGVVAIGMKEETAVSPEEPAKYE
NYPINGVLTAVIEEDGSLGSDVDEPYKEIKKHVKRPVIASGVSLIEDIKKLYGIVEG
VYVGRKAIYEGRTLEDLONP"
complement(11167..11922)
/gene="pyrD8"
/note="aq_1305"
complement(11167..11922)
/gene="pyrD8"
/codon_start=1
/transl_table=11
/product="dihydroorotate dehydrogenase electron transfer
subunit"
/protein_id="AAC07287.1"
/db_xref="GI:2983725"
/translation="MLAEINAEVLENTYISGNLYRLTLRIPDKILKOIBPGHAMIKP
SPYDPMGRAPRAVADIEBKAFYDYDGKGYLSKRRGRLTPIPLDKRLTSY
EGDKHLLIGGGLTGLTLAKLRDMGKVPYATGGRSHGLGYEMLEKEFPYEL
FVNDGSAGRGVTELTIKESTDWVVAAGCPKPLTIEMKTKGKRVYSLERMACG
MGVCLCCVYKTRGCKFRKVCYEGEPVAPMEVIL"
11990..13300
/gene="sucC1"
/note="aq_1306"
11990..13300
/gene="sucC1"
/codon_start=1
/transl_table=11
/product="succinyl-CoA ligase beta subunit"

```

Alignment Scores:

Pred. No.:	Score:	Length:
3.04	144.50	15569
Percent Similarity:	34.18%	Matches: 111
Best local Similarity:	20.68%	Conservative: 111
Query Match:	3.89%	Mismatches: 267
DB:	1	Indels: 277
		Gaps: 40

```

US-10-008-355-2 (1-712) x AE000733 (1-15569)
QY 9 LeuLeuGlyAlaAlaLeu-----LeuLeuGly-----AlaSerGlyValAla 22
||||| ||||||| :|||
DB 3820 CTGCTTACGCGGACCTCAGCAACGTTCTCGGCGACGACGTAAGACAGCGTTCCTC 3879
QY 23 LysAlaAspIysGlyMetIryPLeu-----LeuSngIuLeuAsngIuAsn 38
||||| :|||
DB 3880 GTCGCGGACGAAGTATTGAGGTTTGACTTACACACTTTTCAGCTCTAACGGAAGAGAA 3939
QY 39 LeuAspIryGleAlaGluLeu-----GlyPheMtrIryPLeuAspSer 53
||||| :|||
DB 3940 CTCGAAGAGGGGGAAGAACCTGTGAACGAAATAAGGGAACCTCCCGTGAATGTA 3999
QY 54 Leu---TyrSerPheAspIryProSerIleAlaAsnAlaValIlePheGlyGly 72
::: :|||
DB 4000 ATGGAATGCGATACGATGAAGACATAAAAGCGTGCCATAGCGATATTGAGGAAGAAG 4059
QY 73 -----CysThrGly 75
||| |||

```

Db 4060 TACGGGAAAGGCTGAGCGTAATATCTCGCGAGAGTTTCAAGAACTCTCGGTGGA 4119
 QY 76 ILeThValSerAspGlnGlyLeuIlePheThraSnhHisCysGlyTyrglyAlaIle 95
 Db 4120 ACCGACTTTCGCCGAACTGAGACATTA-----GCCCTACTTCAGATA 4161
 QY 96 GlnSerGlnSerThraVal----- 101
 Db 4162 ATATCCGAGAGCTCCGCGGTGCGGCGTGAAGAAATAGTGCGACAAACGGGAAGGTGG 4221
 QY 102 -----AspHisAspTyLeuArgSpsGlyPheValIserAlaGlyThr 114
 Db 4222 TCCGTAAACACCGCTTCAAGAACACAGACGCTCAAGAACTTCTTCAGCTCTCGGC 4281
 QY 115 MetGlyGluGluLeuProIleProGlyLeuSerValIleTyLeuArgGlySileValIle 134
 Db 4282 GTGGCGCAGAG-----GAAAGTAATCCAG 4305
 QY 135 ValThraSpsIysValGluGlyGlnLeuLysGlyIleThraSpsGlnMetGluArgLeuArg 154
 Db 4306 AAATTTGAAGAACTTAAAGAAATCAAGACAGAGAAAGAAATACAAAGACTGAA 4365
 QY 155 LysAlaGlnIleVal-----CysGlnGlnLeuAlaIleLysGlnSpsGlnAlaAsp 170
 Db 4366 -----CAGGAAGCTCTGAAGCTCCAGATTAAGGAGTTGTAAAGAAAGAAAGCTTGA 4419
 QY 171 GluAsnGlnLeuLysIleValGluProPheTyLeuSerAsnGlnIleTyLeuIleVal 190
 Db 4420 GAT-----TTCAACACTCCAC 4434
 QY 191 TyrAlaValPheLysAspValIleArgMetValPheAlaProPheSerSeraValGlyLysPhe 210
 Db 4435 TACGCGTATTGTGAAGGTTGAA-----CCGAAAGAACTCAGAAACCTC 4479
 QY 211 GlyGlyAspThrSpsAsnTrpMetTrpProArgHisThrGlySpsPheSerValPheArg 230
 Db 4479 ----- 4479
 QY 231 ValTyraIleValAlaAspAsnArgProAlaGluTyLeuSerLysSpsAsnLysProTyLys 250
 Db 4480 -----GCGGATATGCTGAGGCAAGAAAGCAAGAAAGAAC----- 4512
 QY 251 ProValTyLeuPheAlaAlaValIleSerMetGlnGlyTyLeuLysAlaSpsTyLeuAlaMetThr 270
 Db 4513 -----GTTGTGTTTATAGCTCCAGAAAGGAGATTAAGATAACTTCGT 4557
 QY 271 IleGlyIlePhePro-----GlySerThr 277
 Db 4558 ATCGGAGTTTCCAAAGCATATCGACAAAGCTGAACGCCAAGAACTAATTAAGGAGGTC 4617
 QY 277 AspArgTyLeu-----ThrSerTrpGlyValGluAspArg 289
 Db 4618 GGAAGGTACTTTAAGGTGAGAGAGAGAAAGGCAAGCACTCGCTCAGGTGAGAGAAAG 4677
 QY 289 GlnGluAsnGluAsnPro-----ArgIleGluValArgGlyIle 303
 Db 4678 GCTCTCGACA---AATTCGCCGAAGCGTTAAATTACTAAAGAAATCTTAATGCGGT-- 4732
 QY 303 eLysGlnGlyIleTrpLysGlnAlaMetSeraIle-----AspGlnAlaThr 318
 Db 4733 -----AAAGGTGTGATGAAGAAAGACACAGACAGCGGTATCGACTCAAAAGCTGTGAT 4788
 QY 318 ArgIleLysTyLeuSeraIle-----LysTyraIle 327
 Db 4789 AAGGCTGTGAGGAGCTCTCCCTCTCCGTTGGGGGCTTCGTTACTGTAGTAGAGCG 4848
 QY 327 ArgIleSeraIleAsnTyLeuTrpLys-----AsnSerIleGlyMetAsnArgIleVal 344
 Db 4849 AATCCCGCTAAGCTAATCTGAGATAGACGAAAGCAAGTAATAACTGAGATTTCCGTGGA 4908
 QY 344 AsnGluAspValIleGlyTyArgLysArgAlaGluGluValArgAlaPheAlaSpsTrpIleArg 364
 Db 4909 AAAGCTGAATAATTTATCGAGGCTGTGTGAATAACATAGAACAAATTGTGTGACAAAGTTGT 4968

QY 364 GlyAsnGlnLysSerAlaValTyrglyLysPValLeuSerSerLeuGlnLysAlaTyrgly 384
 Db 4969 AGAAGAGTTGAGGTAGCCCTCGCCGAGGGTGTTTC---ACATATAGCAAGATTATACCG 5025
 QY 384 SGIuGlyAlaLysAlaAsnArgGlnMetThrTyLeuSerGlnIlePheGlyLysIle 404
 Db 5026 AAACGGTATTTAAATC---AGAAAGCTCTTACTTGAAGAG----- 5065
 QY 404 rGluValValArgPheAlaGlnPheAlaAsnAlaLeuAlaIleAsnProAspAlaHisAl 424
 Db 5066 -----TACACTTCCGTGGGTAAAGAAATATATATAGACAGAGAC----- 5104
 QY 424 ArgIleLeuLysSerLeuAspSpsTyrglyAspTyLeuProSerLeuAspArgly 444
 Db 5105 -----AGCGAACTGATGAGAAATCCCGCTTACGTTAAGAAACAGGGAAC 5154
 QY 444 SValLeuProAlaMetLeuAspIleValIleArgArgGlnIleProAlaSpsLysLeuProAs 464
 Db 5155 GATCTCTCGCACATATAGGAGAACTCATCTGCTGAGCTCATGTACTTCTTAAAGAA 5214
 QY 464 PhePheLysAsnValIleAspLysLysPheLysGlyAspThrLysLysTyraIleAspThr 484
 Db 5215 CATTAATTAAG---ATATATACACTGTTTGTATAGCCCTTGAAAAACCTTAGCTTTT 5271
 QY 484 eValPheAspLysSerValValProTyLeuSerAspLysPheHisAlaMetLeuLysSerIle 504
 Db 5272 TTTATTTTATTAATAAACTGAGAGAGTAAAGACATGTT----- 5311
 QY 504 LAspLysGlnLysPheAlaLysAlaIleGluLysAspProAlaValGluLeuSerLysSe 524
 Db 5312 -----GAGAAAGTTC-----ACGAAAGGCGCAAGACA 5337
 QY 524 rValIleAlaAlaIleArgAlaIleGlnIleAlaAspAlaMetAlaAsnAlaTyraIleG 544
 Db 5338 AGTTATCTTGAGGCGGAGG-----GAAAGAGCCCTCGAGCTCGGACACAGCTACTT 5388
 QY 544 uLysGlyLysArgLeuPhePheAlaGlyLeuArgGlnMetTyLeuProGlyArgAlaLeuPr 564
 Db 5389 GGGAACTGAACATATTTTGTGCTGCTGATTAAGATGAGGA---CATACCCGTCTGCTGC 5447
 QY 564 oSerAspAlaAsnPheThrMetArgMetSerTyrglySerIleLysGlyTyrglu--Pr 583
 Db 5448 TCAGCAAGCTGGGTGT-----CTCTCTGAGAAAGATTAAGAAAGGCAATATAGGCG 5498
 QY 583 oGlnAsp-----GlyAlaTrpTyraSnyrHisThrThrGlyLysGly-- 597
 Db 5499 AATTAATTAAGGAGACCCACTCGGAGAGG---TACTCTTCGCCGCGACCAAAAGGTA 5557
 QY 597 ----- 597
 Db 5558 ATAGATTTGCAAGTTGAAGAGGAGAAATTTCTACACACAGTTGCTAGTCCGGAACAC 5617
 QY 598 -----ValLeuGlnLysGlnAspProLysSerAspGlnPheAlaValGlnI 613
 Db 5618 CTACTTATCGAGTGTGTAGAAAAGAGCGGCACTCGTGAAGAAATTT-----TA 5668
 QY 613 uAsnIleLeuAspLeuPheArgThr-----LysAsnTyrglyTyraIleGlnLys 630
 Db 5669 AGGAGCTTTGGAGTGGACAGATCTCCGTAAGAGAGAGACTTCTCACTACTCGAGAGA 5728
 QY 630 nGlyGlnLeuHisIleAlaPheLeuSerAsnAspIleThrGlyLysSerGlyLys 650
 Db 5729 CTTCGCCCTCAAGAAAGCC-----AGAGGTATCGTA---TCCAC 5763
 QY 650 rProValPheAspLysAsnGlyArg---LeuIleGlyLeuAlaPheAspGlyAsnTrpI 669
 Db 5764 TCCTTAACCTTTCAGAGCTTCTCAAGGACCTCAACCCAAAGGCAAGGAGAAAGCTTCA 5823
 QY 669 uAlaMetSerGlyAspIleGluPheGluProAspLysGlnIleArgThrIleSerValAspI 689
 Db 5824 CCCAGTAATAGCGAG-----GAAAGGAAATATAGAGAGATTTATCCAGATTACTCGT 5874

QY 689 earg 690
 Db 5875 ANGA 5878

RESULT 5
 OSJN00169
 LOCUS Oryza sativa chromosome 4 clone OSJNba0043124, *** SEQUENCING IN
 DEFINITION PROGRESS *** in ordered pieces.

ACCESSION AL662969
 VERSION AL662969.1 GI:17998481
 KEYWORDS HTG; HTGS_PHASE2.
 SOURCE Oryza sativa.
 ORGANISM Oryza sativa
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzeae; Oryza.

REFERENCE
 1 (sites)
 Han, B., Feng, Q., Huang, Y.C., Li, Y., Zhu, J.J., Zhao, Q., Hu, X.,
 Liu, Y.L., Mu, J., Yu, Z., Chen, L., Fan, D.L., Meng, Q.J., Zhang, L.,
 Lu, X.Q., Yu, S.L., Liu, X.H., Lu, T.T., Zhang, X.J., Lu, Y., Li, C.,
 Li, T., Zhang, Y., Hu, H., Jia, P.X., Qian, Y.M., Yang, K., Zhou, B.,
 Chen, Z.H., Hao, P., Zhang, L., Wu, M., Zhang, R.Q., Guan, J.P., Fu, G.,
 Wang, S.Y., Ren, S.X., Lv, G., Liu, W., Gu, W.Q., Zhu, G.F., Tu, Y.F.,
 Jia, J., Yin, H.F., Zhang, Y., Cai, Z., Chen, J., Kang, H., Chen, X.Y.,
 Shao, C.Y., Sun, Y., Hu, Q.P., Zhang, X.L., Zhang, W., Wang, L.J.,
 Ding, C.W., Sheng, H.H., Gu, J.L., Chen, S.T., Ni, L., Zhu, F.H. and
 Hong, G.F.
 Direct Submission
 Submitted (27-DEC-2001) Han Bin, National Center for Gene Research,
 Chinese Academy of sciences, 500# Cao Bao Road, Shanghai 200233,
 CHINA. E-mail enquiries: bhan@ncgr.ac.cn. Clone requests:
 bhan@ncgr.ac.cn
 Oryza sativa japonica (nipponbare) genomic DNA, chromosome 4, BAC
 clone: OSJNba0043124.
 Web site: http://www.ncgr.ac.cn
 ----- Summary Statistics
 Assembly program: phrap

 * NOTE: This is a 'working draft' sequence.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 Location/Qualifiers
 1.162595
 /organism="Oryza sativa"
 /variety="Nipponbare"
 /sub_species="japonica"
 /db_xref="taxon:4530"
 /chromosome="4"
 /clone="OSJNba0043124"
 /clone_11b="CUG1-OSJNba"
 BASE COUNT 44942 a 35736 c 36400 g 45514 t 3 others
 ORIGIN

Alignment Scores:
 Pred. No.: 168 Length: 162595
 Score: 137.50 Matches: 147
 Percent Similarity: 36.64% Conservative: 115
 Best Local Similarity: 20.56% Mismatches: 260
 Query Match: 3.70% Indels: 195
 DB: 2 Gaps: 34

US-10-008-355-2 (1-712) x OSJN00169 (1-162595)

QY 47 PheThrLeuProLeuAspSerLeuTyrSerPheAspLysProSerIleAlaAsnAlaVal 66
 Db 108215 TTGTGTCAGCGCTCTGCTTCTTA-----TCACATGCAAGTGTGTACA 108256

QY 67 ValIlePheGlyIleGlyCysThrGlyIleThrValSerAspSngIleuIlePheThr 86
 Db 108257 AGTCACCTTGATGATGTCAGCTCATGTGCTATGACAA----- 108301

QY 87 AsnHisHisCysGlyTyrGlyAlaIleGlnSerGlnSerThrValAspHisAspTyrLeu 106
 Db 108302 -----ACTGCTGCTATTCTTCAAGTCACTCTTTTCATCAATACTCTT 108346

QY 107 ArgAspGlyPheValSerArgThrMetGlyGlnGluLeuProIleProGlyLeuSerVal 126
 Db 108347 CGTTTGTCA---CAGTCANAAGCTGTGGAGAGAG-----TCTGAA 108385

QY 127 LysTyrLeuArgLysIleValLysValThrAspLys-----ValGlnGlyGln 142
 Db 108386 AGAAGCCCTAGAGAGAGTGTCTTCACTCAAGCAAGAAAGGTTTCTGCCATTGAAAGTTTG 108445

QY 143 LeuLysGlyIleThrAspGlnMetGluArgLeuArgLysAlaGlnGluValCysGlnGlu 162
 Db 108446 CTGAAGGCTGTAGC----- 108460

QY 163 LeuAlaLysLysGlnAsnAlaAspGlnAsnGlnLeuCysIleValGlnProPheTyrSer 182
 Db 108461 ATATCAGATAGAGCAAAATATCTCAGCTACCGCTCAACAGCTGGATCT-TGGTATCTCT 108519

QY 183 AsnAsp-----GluTyrPheLeuIleVal-----TyrAspValPhe 194
 Db 108520 TTTAACTGCTTAGTGTCTCAATTACTGTTAAGTAAATGGAATGCGCTTTATCATGTGATTT 108579

QY 195 LysAsp-----ValArgMetValPheAlaProProSerSerValGlyLysPheGlyGly 212
 Db 108580 CCTGGTCATTGACTGCAATATATCATCTTACTTTTCTATGATTAATAAATC----- 108633

QY 213 AspThrAspAsnThrMetThrProArgHisThrGlyAspPheSerValPheArgValTyr 232
 Db 108634 -----AGTTCTTATTCTCT-----TTTAATCTTAGT 108660

QY 233 AlaGlyAlaAspAsnArgProAlaGluTyrSerLysAspAsnLysProTyrLysProVal 252
 Db 108661 GCAGGGGTTGAT-----CCTCATCGCTCTCGAT-----CCTCTGT 108699

QY 253 TyrPheAlaAlaValSerMetGlnGlyTyrLysAlaAspTyrAlaMetThrIleGly 272
 Db 108700 CCGCTTGCAACACAGCTTCAATCATCTGTCTTACAGAAATTCAGACTGCTCGACTCA 108759

QY 273 PheProGlySerThr-----AspArg 279
 Db 108760 TCTGTTCCTAGACACCAATATGCTAGTGCAGAAATGAGAGCTCCGCTATTGAGTCA 108819

QY 280 TyrLeuThrSerTyrGlyValGluAspArg-----IleGlu 291
 Db 108820 ATGACAACCCAGTGTGGCGACAGAGCATCAAGATCACCGTACTTGGTATATATCA 108879

QY 292 AsnGluAsnAsnProArgIleGluVal-----ArgGlyIleLysGlnGlyIle 307
 Db 108880 TCTGATCTCATGACAAAGCTTATCATCTGCTTTCCAAAGAGGTCTTTGGAGAGGCCA--- 108936

QY 308 TrpLysGluAlaMetSerAlaAspGlnAlaThrArgIleLysTyrAlaSerLysTyrAla 327
 Db 108937 ---CAAGAAGGAGCGCGCATGATGATGAGGATGATATCCGTCACACTAGAGATTTCC 108993

QY 328 GlnSerAlaAsnTyr-----TrpLysAsnSerIle----- 337
 Db 108994 CAACACTCAAAAGCTTTTGCATCATGCTTACAGAGATGCCATCCAGAGATTCACATTAAT 109053

QY 338 -----GlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLysArg 353
 Db 109054 ATCATGTTCCAAATTTCCAGAGACCTCTTCAAGAGACCAAGTAAATGTCAGAGGCTTCT 109113

QY 354 AlaGlnGluArg---AlaPheAlaSerTyrPheArgLysAsnGlyLysSerAlaValTyr 372
 Db 109114 GCCAGATCAGCAGCAGCTTTGACATAGCAGTACCAAGTACAGTCAAGCTGCTGTGTTAT 109173

QY 373 GlyAspValLeuSerSerLeuGlnLysAlaTyrLysGlnGlyAlaLysAlaAsnArgGlu 392
 Db 109174 ACGGATGACACTGCTTCACTAAGTATGATGCTTTTGCAGAGGT----- 109215

QY 393 MetThrTyrLeuSerLcuthrLeuPheGlyGlyThrGlnValValArgPheAlaGlnPhe 412

```

Db 109216 -----CTCAGCCCTGATGTCGGATG-GTACTGAGAGCTCAGCTTTGAATTAT 109265
Oy 413 AlaAsnAlaLeuAlaThrAsnProAspAlaHisAlaGly-IleLeuLysSerLeuAspAs 432
Db 109266 TCGGAATTTTCTGCAACAGAGGATATTCAGAAATTCACAGAAATTTGAA 109325
Oy 432 PLSyTyrLysAspTyrLeuProSerLeuAsp-----ArgLysValIleProAlaMe 449
Db 109326 GGTCAATGAAGCGTTTTCGCACTTGTGATGACCCCATCATTAAGTTTGCACAAGCAGC 109385
Oy 449 tLeuAspIleValAlaArgArgAlaGleProAlaAspLys----- 461
Db 109386 CTTTCTACTCTCTTCAGAACTATATTCAGCTTCAGAGAAGCCATTGAGATTATGTTGA 109445
Oy 462 -----LeuProAspIlePheLysAsnValIleAspLysLysPheLysGlyAspThrLys 479
Db 109446 AAGAAATTTTGCCATATGTTTTTCCCGGCTTATGTGATCCAAAGAAATG-----GTAAA 109499
Oy 479 sLysTyrAlaAspPheValIleAspLysSerValValProTyrSerAspLys-----Ph 497
Db 109500 AAACCATGCGTCATCAACCTTGATGTTGTTGGAGAAACATTCGAATTGATGTGCT 109559
Oy 497 eHisAlaMetLeuLysSerMetAspLysGlyLysPheAlaLysAlaIleGlyLysAspPr 517
Db 109560 ACCTCGCTAGCGCTCAGATGATGAACAGAGTCCCAAGGCA-----AAGTTGCG 109613
Oy 517 oAlaValGleLysSerLysSerValIleAlaAlaAlaArgAlaIleGlnAlaAspAlaMe 537
Db 109614 TGTCTTGAAGTTGCGCAAAATTAATGCTGTC--AGCAAGTACACTGTAAATTCGAAGGTTA 109670
Oy 537 tAlaAsnAlaTyrAlaIleGlu----- 544
Db 109671 TAGTAAACAGTGCGCTCTTAAACGTGGCTTTCAAAAATTCGCACTTATGTTATCATGAAAA 109730
Oy 545 -----LysGlyLysAlaGlyLeuPhePheAlaGlyLeuArgGluMetTyrProGlyArgAl 562
Db 109731 AAATGCAAAAGTGAAGGAGGAGCATTCATTTCTGATCATATCTGTTATTCATTTGCA 109790
Oy 562 aLeuProSerAlaPheAlaPheMetArgMetSerTyr----- 575
Db 109791 TTCAACACAGAGCTAAATTTTATTTCTAACTATCTGTTGAGAGCAAAACCTTTGAG 109850
Oy 576 -GlySerIleLysGlyTyrGluProGln--AspGlyAlaTyrTyrAsnTyrHisThrTh 594
Db 109851 GCGTCACTGAAGCAATGACACCTCGCATTCGATTAATTTAGTAAGTACTTACAGAG 109910
Oy 594 rGlyLysGlyValLeuGluLysGlnAspProLysSerAspGluPheAlaValGlnGluAs 614
Db 109911 CAAGAAA-----GATCGTCCCGTCCAAATCTTATGATCAAGCA----- 109950
Oy 614 nIleLeuAspLeuPheArgThrLysAsnTyrGlyArgTyrAlaGluAsnGlyGlnLeuH1 634
Db 109951 -----GATTATGCAACTCTTCGCAAGATGG----- 109977
Oy 634 sIleAlaPheLeuSerAsnAsnAspIleThrGlyLysAsnSerGlySerProValPheAs 654
Db 109978 -TACCGCTTGCTCAAGAAAGCTATTCATTCATTCGCGGATTCCTAGTCCCTTGA 110036
Oy 654 pLysAsnGlyArgLeuIleGlyLeuAlaPheAspLysAsnTrp 668
Db 110037 TGCTGAAGT-----GGGAAATGG 110055

```

```

REFERENCE 1 (bases 1 to 4004)
AUTHORS Kimura,T. and Horikoshi,K.
TITLE The nucleotide sequence of an alpha-amylase gene from an
JOURNAL alkaloisymbiotic Micrococcus sp
FEATRES FEWS Microbiol. Lett. 71, 35-42 (1990)
source location/Qualifiers
1..4004
/organism="Micrococcus sp."
/strain="207"
/db_xref="taxon:1271"
155..189
/rept_type=INVERTED
265..272
RBS 282..3596
CDS
repeat_unit
155..189
/rept_type=INVERTED
265..272
RBS 282..3596
CDS
/transl_table=1
/product="alpha-amylase"
/protein_id="CAA59321.1"
/db_xref="GI:296763"
/translation="MSKKMKOKFLMVLALLVIGDAVCSATGAQETSPADREVR
LSTIVRINDDMIFGRHRRRLMKSTLSRMKGVRSRHVAAPRNREGIWFCAOL
KLDSFPRKINGSLLQVKKDSLQAFVTSGETQIRIVPDGAPVIDNGENMYRDR
LFADADAMNTLEKELDFDGETHEMYEPENERVFLANDIPTEBHPYTLVTAQVT
EVDIPNTVGVDSILVFNQADLAVSGTAVENALDFNONAVSDIEGITEDVRSKIT
ADTSLGSAOLEIDPTLKEISLAVDDIAGAKALPLVTYDSGINKSTAVENVT
ROSYGEGDFPDMDSVITYFMLTDFPDQDANNDPNDYDNKESRGITGGDFGIDPK
LDYIDELGVNTISPVENIKKDYRERYESPEYCYGHCYMANFGLNHPTEMEF
HDILDGADHDMKIMVDVNVNHTGYLKEIDGYSVTPNPACTPSDADARFSDLROCA
DVGTEDEVGELAGLPDITEDPNVRKQIDMOTDIMEKATENGNTIDYRVTYKHV
EDATWMOFKNALTEKMEPEKMEGAMCAKVDNTLGYLETGMSLDLDFPKETARSV
NGSLAANAISLTARNAKLDTATLGOFLSHDEEGFLSLADGKGLQVAAATLQATK
GOPYVITYGSELGQGANNPYOYRNDYFANDVEGNELIAYTKILFREGYSKVEAK
GERTVGSGDKQDFLESRDYOKYVYVGLNVAESKAVTLTVDSADVAVTDAVSGTE
YRTACGVNNTLTPKADPTVLTATVEGNTITCAKDNQERYVLEPNNIRIHKRED
NVYKNGAWLMDNVASPSANVPVATVEKTDSTGATIDVLDGANNIGFLVMDITA
GDAGKDDGDFGTISSPOANEIYKQSDRVYVEPVLDPAANTVRIHYTREADVDF
GIWNMGDVASPDGMPTGAVNFQDIDRYGAYPIKLEAKKEIGFLVNNKTOEKDAG
DKSFNLDKFNHMLIKQDDTVYVSPDEYVASGLGSAEVVSEMTLLRFNSTGTGTFSE
KIVAGLAIQDADSAAL"
BASE COUNT 1239 a 724 c 1028 g 1013 t
ORIGIN
Alignment Scores:
Pred. No.: 3.45 Length: 4004
Score: 133.50 Matches: 138
Percent Similarity: 34.94% Mismatch: 259
Best Local Similarity: 20.78% Indels: 173
Query Match: 3.59% Gaps: 35
DB: 1
US-10-008-355-2 (1-712) x MSP207AAM (1-4004)
Oy 141 GlyGlnLeuLysGlyIleThrAspGluMetLuarGleuAlaGlyAlaGlnGluValCys 160
Db 141 GCGCATTTTAAAGCATTAACAGCAACGCTGATTAATTA----- 1469
Oy 161 GlnGluLeuAlaLysGlyAsnAlaAspGluAsnGlnLeuCysIleValGluProPhe 180
Db 1470 -----GATGACCTTGCGGTTAATACAAATTGGATT--AGTCCGGTT 1508
Oy 181 TyrSerAsnAsnGluTyrPheLeuIleValTyrAspValPheLysAspValArgMetVal 200
Db 1509 GTGGAACAACATTAAAGTAC-----GAGTTCGCTTATTAAC 1541
Oy 201 PheAlaProProSerSerValGlyPheGlyGlyAspThrAspAsnTrp-----Met 218
Db 1542 GAACGTCAGAACCGTATTAATCACTGATCGGCAAAATTAATTTTGGAGAGTTA 1601
Oy 219 TrpProArgHisThrGlyAspPheSerValPheArg--ValTyrAlaGlyAlaAspAsn 237
Db 1602 AATCGC---CATTTTGGCAGCATGGAAGAGTTTCAATGACTTGATGACGGGGCACACAGAC 1658

```

QY	238	ArgproAlaGluTyrSerLysAspAsnLysProTyrLysProValTyrPheAlaIaVal	257
		:::	:::
Db	1659	CGCAACACGAAAGAAATATGATGTCGAT	GTGGTCGTC 1691
QY	258	SerMetGInGlyTyr-----LysAlaAspAspTyrAlaMetThr-----	IlleGly 272
		:::	:::
Db	1652	AATCACACAGGATACGGCTTTAAAGAAATGATGGTTGGTCACTAATCCGCCACAGCGGC	1751
QY	273	PheProGlySerThrAspArgTyrLeuThrSer-----	283
		:::	:::
Db	1752	TATCGTCAGATGCCGACCGGCAAGTTTAGTGATTTGCTTCCTCGTCAAGGGCGGACGTA	1811
QY	284	-----TyrGlyValGluAspArgIleGluAsnGluAsn	294
Db	1812	GGTACAGATGAAGCGTGGTGATTAATACGGCGGACTTCCAGCACTTATACGGAAATCCA	1871
QY	295	AsnProArgIleGluValaArgGlyIleLysGInGlyIleTyrPlySGluAlaMetSerAla	314
		:::	:::
Db	1872	AACGTTCGCAACCAAAATATCAT--TGCAAAACCGATGGATTGMAAAACGCCAGCA	1928
QY	315	AspGlnAlaThrArgIleLysTyr-----AlaSerTyrValaIleSerAla	330
		:::	:::
Db	1929	GAAGACGCATACGATTTGATTTATTCCTCGTGTGATACGCTGAACACCTACAGATCCG	1988
QY	331	AsnTyrTyrPlyAsnSerIleGlyMetAsnArgGlyLeuAlaArgLeuAspValIleGly	350
		:::	:::
		:::	:::
Db	1989	ACGTGATGACAGTTTAAAAATGCATTAACGGAAAAATCCGGAATTTTAAATGATCGT	2048
QY	351	ArgLysArgIaGluGluArgAlaPheAlaAspTyrIleArgLysAsnGlyLysSerAla	370
Db	2049	GAAGCT-----TGGCGCGCTAAAGTGGACATCTTTA	2081
QY	371	ValTyr-----GlyAspValLeuSerSerLeuGluLysAlaTyrTyrGluLysIle	387
		:::	:::
Db	2082	GGCATTCTTGAACACGGCACAGATGATTCATTCTTGACTTTGGTTTAAAGAAACGGCT	2141
QY	388	Lys-----AlaAsnArgGluMetThrTyrLeuSerGlu	398
		:::	:::
		:::	:::
Db	2142	CGGTCATTTGTTAACGGCAGTTTAAAGAACGGCGTAATAGCCTCTTACTACACGAACCG	2201
QY	399	ThrLeuPheGlyGlyThrGluValaValArgPhe-----AlaGlnPhe	412
		:::	:::
Db	2202	AAATGGATTAATACGCAACACTCGGCAAGCTTTTGAAGAACCATACGAAAGAGGTTCC	2261
QY	413	AlaAsnAlaLeuAlaThrAsn-----	419
Db	2262	CTTCATTCACTTGTTGGGGAATAAAGAAATTAACAAGTCGACGCCACTTTCACACACA	2321
QY	420	-----ProAspAlaHisAlaGlyIleLeuLysSerLeuAspAspTyrLys	435
		:::	:::
Db	2322	GCMAAAGCACACCTCGTATTTACTACGCA-----GAAGAGCTTGGCCAAACAGGACT	2375
QY	436	AspTyrLeuProSerLeuAspArgLysValLeuProAlaMetLeuAspIleValaArgArg	455
		:::	:::
Db	2376	AATATTATTCACAATACGATTAATCGT-----TAGATTTTCCGGGAT	2420
QY	456	ArgIleProAlaAspLysLeuProAspIlePheLysAsnValIleAspLysLysPheLys	475
		:::	:::
Db	2421	CAAGGCGAAGCAATGAAATTTTGGCCATTAACCCMAATTTTAAACTTTAGA--GAA	2477
QY	476	GlyAspThrLysLysTyrAlaAspPheValaPheAspLysSerValaValProThrSerAsp	495
		:::	:::
Db	2478	GGTTTTCGAAAGCTTTGGCAAA-----GGTGAACGCAAGCTTGTTGGAGGATCTAT	2531
QY	496	LysPheHisAlaMetLeuLysSerMetAspLysGluLysPheAlaLysAlaIleGluLys	515
		:::	:::
Db	2532	AAAGATCAATCTCTTATTTTTCGCGACTACATCAAGATCAAAAAGTACCTTGGGTTA	2591
QY	516	AspProAlaValaGluLeuSerLysSerVal-----IleAlaAlaAlaArgAlaIle	532
		:::	:::
Db	2592	AACGTGAACGAAGAA--ACTGAAGCGGTGACTTTAACCGTGGATTTCCGACACGCTCTT	2648

QY	533	GIN1A1aSP1aMetAlaAsnA1a---	TyrAla1a1leGluLysGlyLysArgLeuPhe	551
Db	2649	GTACCGAAGTACCTATCGGGTACGGAAATATACGCAACAGCAGGAAAA		2696
QY	552	AlaGlyLeuArgGluMetTyrProGlyValArgAlaLeuProSerAspAlaAsnPheThrMet		571
Db	2697	-----GTAATATTATACCTCTTCAGAGAAAAGCATCCCGGACCGCTTTATTGACG---		2747
QY	572	ArgMetSerTyrGlySer11leYsgly		580
Db	2748	---GTAGAAAGGAGGACACATATACCGGTGTGGCGAAGAACACACCGGTGAAGTGGTTGAA		2804
QY	581	-----	-----TyrGluProGlnAspGlyAlaTrpTyrAsn	590
Db	2805	CTAGTACCGGAAACAAATATCCGATTCATCATATTAAGCCGCAAGACAAATGTATATAAAC		2864
QY	591	TyrHisThrThrcLysGlyValLeuGluLysGlnAspProLysSerAspGluPheAla		610
Db	2865	TAC-----GGGCGATGGCTATGGAATGATGTTGGCTCTCTCTTCGCAAACTGGCCA		2915
QY	611	ValGlnGluAsn11leLeuAspLeuPheArgThrLysAsnTyrGlyValArgTyrAlaGlu---		629
Db	2916	GTAAGGGGACACATTTTCGAG-----AAACCGCATAGTTACGGCGCTATATCATGTG		2969
QY	630	-----AsnGlyGluLeuHis11leAlaPheLeuSerAsnAsp11leThrcLysGly		646
Db	2970	CCATTAGCTAGCGAGCGCAAAATATTCGGCTTTCTTGTTC---ATGATATTAACCCGGGCT		3026
QY	647	AsnSerGlySerProValPheAspLysAsnGlyValGlyLeu11leGlyLeuAlaPheAspGly		666
Db	3027	GACGGGGGCAAGACGCGCGAGATTAAGCGTTTACGATTTCTTCACCAACGAACAAATGA		3086
QY	667	AsnTrpGluAlaMetSerGlyAsp-----11leGluPheGluPro---AspLeuGlnArg		683
Db	3087	ATTGTGATCAACCAAGCTTCGCAACAAGTGTACAGTACAGCAACCGGTGATTGGCAGCA		3146
QY	684	Thr11leSerValAsp11leArgTyrValLeuPheMet11leAsp-----		697
Db	3147	AAT---ACAGTCCGATCATTCATTAACACGGGTGAAGCGGTATTAAGATGATCTTGGCATPA		3203
QY	698	---LysTrpGly 700		
Db	3204	TGGAATTTGGGCG 3215		
RESULT 7				
LOCUS	AE008428	10292 bp	DNA	linear
DEFINITION	Streptococcus pneumoniae R6 section 44 of 184 of the complete genome.			BCT 13-SEP-2001
ACCESSION	AE008428	AE007317		
VERSION	AE008428.1	GI:15458055		
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				


```

gene      7774..8544
           /gene="spr0493"
CDS       7774..8544
           /gene="spr0493"
           /codon_start=1
           /transl_table=11
           /label="spr0493"
           /product="Conserved hypothetical protein"
           /protein_id="AAK9297.1"
           /db_xref="GI:15458064"
           /translation="MMKVMKSHPHYELNLIGYLAKEFKDLFIKEFQSSKSEFYR
           VVSIQIAETGVKNNRMDLEDFEDNNRKKMMOKAEVYFRRKLDIDMFGENEHSYA
           ETIKMLASEGKKGTITIEVKPIVTRFKRIOTKEAEANFYELIHDEKKEKGGOLT
           DARLYGDGDDVDNOEYSLAEVNGIKRSGKGVRTLAEPFVKRFQSDFILSLVTN
           LDDIPKLYIIDNPKHFERKNITIKETIERSVEDLY"
gene      8827..9630
           /gene="spr0494"
CDS       8827..9630
           /gene="spr0494"
           /codon_start=1
           /transl_table=11
           /label="spr0494"
           /product="Hypothetical protein"
           /protein_id="AAK9298.1"
           /db_xref="GI:15458065"
           /translation="MQPYNIIDNPNTSYAKRDMRIGRGLOKVDNLVSAYESLAE
           KOSRGLTYEYEDATYAHHTIDASTERADLVSLIKVELSRHFSFSPATILAIHK

Alignment Scores:
Pred. No.:      11.9      Length:      10292
Score:          133.00     Matches:      106
Percent Similarity: 33.20%  Conservative: 64
Best Local Similarity: 20.70%  Mismatches: 164
Query Match:      3.58%     Indels:      178
DB:              1        Gaps:      25

US-10-008-355-2 (1-712) x AE008428 (1-10292)
QY 192 AspyAlphe-----LysAspyAlArgmEtValPheAlaPro 203
    |||||
Db 5143 GATGTTTCAAGCCTTGAGCGATCAAAAGCGTAAGCCTTATTCACGTTATTCACCA 5202
QY 204 ProserSerValGlyLysPheGlyAspThrAspAsnTrpMetTrpProArgHisThr 223
    |||:::
Db 5203 CCAAACTTACAGGTAACTT-----CACCTT 5229
QY 224 GlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgProAlaGluTyrSer 243
    |||
Db 5230 GGT-----CACGCT 5238
QY 244 LysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGinglyTyrLys 263
    |||
Db 5239 TGGGATTCAC---ACTTGGACAGATATATATCATCCGTCAAAACGCAACGCAAGGTTTGAT 5295
QY 264 AlaAspAspyTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrLeu---Thr 282
    |||:::
Db 5296 -----ACCTTTGGCTTCCTGGATGGATGCACGACGACGAGGATTCGCACCT 5337
QY 283 SerTrpGlyValGluAspArgIleGluAsnGluAsnAsnProArgIleGluValArgGly 302
    |||||
Db 5338 CAGCGTAAAGGTTGAGAGCCCTTCGCTGGAGGCAATTCCCGCTATGACCTTGTCGT 5397
QY 303 -----IleLysGinglyIleTrpLysGluAlaMetSerAlaAspGlnAlaThr 318
    |||
Db 5398 GAGTCTTTTTCACGAAAGCTGGGAATGGAAGAC-----GATATATGCCACT 5445
QY 319 ArgIleLysTyrAla---SerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsnSerIle 337
    |||||
Db 5446 ACCATCAAGACAAATGGGCAAGATGGGCTCTCTGTACACTTCTCTCGAGACCTTTC 5505
QY 338 GlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGluArg 357
    :::::
Db 5506 ACTCTTGACGAAGGTTTGTC-----AAAGCTGTCGTAAG 5541

```

```

QY 358 AlaPheAlaAspTrpIleArgLys-----AsnGlyLysSerAlaValTyrGly 373
    |||
Db 5542 GTCTTTGTGACCTTACAAAGAAAGGCTGATCTACCGTGATGGATTATCATCACTCG 5601
QY 374 Asp-----ValLeuSerSerLeuGluLysAlaTyrLys----- 384
    |||
Db 5602 GACCCAGACACTGGCACAGCCCTTTCTGATATTTAGAGTGATTCACAAAGATGTGGAAGCT 5661
QY 385 -----GluGlyAlaLysAlaAsnArgGluMetThr 394
    |||
Db 5662 GCCTTACACACATGATATCATAGCTGTGAGAGATGGTTCACGGCCCTTGAAGTTGTACAA 5721
QY 395 TyrLeuSerLurThrLeuPheGlyGlyThrGluValAlaArgPheAlaGlnPheAlaAsn 414
    |||||
Db 5722 ACTGTCCTGACACTAGTATGTTTGGGAGCT----- 5751
QY 415 AlaLeuAlaThrAsnProAspAlaHisAlaGlyIleLeuLysSerLeuAspAspLysTyr 434
    |||
Db 5752 GCGATTGCGGTTAATCCAGAA-----GACCGCGCTAC 5784
QY 435 LysAsp-----TyrLeuProSerLeuAspArgLysValLeuProAla 448
    |||||
Db 5785 AAGCACTTGATGTGTAAAAACGTCATCTTCCA---ATCGCTAATAAATCATCCCAATC 5841
QY 449 MetLeuAsp-----IleValArgArg 455
    |||
Db 5842 GTTGGAGATGACGACGAGATCTGAGTTGCTAGTGTGCTGTAAGAAACACACCTGCC 5901
QY 456 ArgIleProAlaLysPheLysLeuProAspIlePheLysAsnValIleAspLysPheLys 475
    |||
Db 5902 CAGCATTCAAATGACTCTTGTGGTGGCCACGTCATATACCTTCCACAAAGTCACATCATG 5961
QY 476 GlyAspThrLysLysTyrAlaAspPheValPheAspLysSerValValProTyrSerAsp 495
    |||
Db 5962 AACGACGACGACACTAGTAAACGACGCTGCTTGAATTTACAGGACATG-----GAT 6012
QY 496 LysPheHisAlaMetLeuLysSerMetAspLysGluLysPheAlaLysAlaIleGluLys 515
    |||
Db 6013 CGTTTGAACCTCGT-----AAGCAGCTGCTGTCTAAGTTGGAAGAAATC 6057
QY 516 AspProAlaValGluLeuSerLysSerValIleAlaAlaAlaArgAla----- 531
    |||
Db 6058 GGTGCCCTGTCAAAATCGAAAAACGTCTCACAGTGTGTGCTACTCAGAACGTCACAGCT 6117
QY 532 -----IleGlnAla 534
    |||
Db 6118 GTGCTACTGAGCCACGCTGTGTACTCATATGTTGCTCAAGATGACCAATTGGCTAAG 6177
QY 535 AspAlaMetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPhePheAlaGlyLeu 554
    |||
Db 6178 AACCCCATTCGCAACCAAGACACAGAGGACAAAG----- 6210
QY 555 ArgGluMetTyrProGlyAlaGlnAlaLeuProSer----- 565
    |||
Db 6211 GTGCAATTTCAACACCTCGTTTCAACGATACCTTCTTCAATGATGAGAAATGTCCAC 6270
QY 566 AspAlaAsnPheThrMetArgMetSerTyrGlySerIleLysGlyTyrGluProGlnAsp 585
    |||
Db 6271 GACGCGGTATATCTCTGTCAGCTCTGCTGGGGCTCAC-----CAATC 6312
QY 586 GlyAlaTrpTyrAsnTyrHisThrThrGlyLysGlyValLeuGluLysGlnAspProLys 605
    |||
Db 6313 CTGCTTGTGTCATAT-----GCTGATGTGTAAATGTATGTGCGGCAAGAACTCCAGAA 6366
QY 606 SerAspGluPheAlaValGlnGluAsnIleLeuAsp 617
    |||
Db 6367 GGTGACGAGATGACCTACGACGAAGAGCTTGTGAT 6402

```

RESULT 8
 SPNEU1923/c SPNEU1923 11362 bp DNA linear HTG 11-JUL-2001
 LOCUS Streptococcus pneumoniae clone G54, *** SPOUENCING IN PROGRESS ***
 DEFINITION

in ordered pieces.
 ACCESSION AL449945
 VERSION AL449945.1 GI:11545170
 KEYWORDS HTGS_PHASE2.
 SOURCE Streptococcus pneumoniae.
 ORGANISM Streptococcus pneumoniae
 Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae; Streptococcus.
 REFERENCE 1 (bases 1 to 11362)
 AUTHORS Dopazo,J., Mendoza,A., Herrero,J., Caldera,F., Humbert,Y., Friedli,L., Guerrier,M., Grand-Schenk,E., Gandin,C., de Francesco,M., Pollisi,A., Buell,G., Feger,G., Garcia,E., Peltsch,M. and Garcia-Bustos,J.F.
 TITLE Annotated draft genomic sequence from a Streptococcus pneumoniae type 19F clinical isolate
 JOURNAL Microb. Drug Resist. 7 (2), 99-125 (2001)
 MEDLINE 21353329
 REFERENCE 2 (bases 1 to 11362)
 AUTHORS Dopazo,J., Mendoza,A., Herrero,J., Caldera,F., Pollisi,A., Humbert,Y., Friedli,L., Guerrier,M., Grand-Schenk,E., Gandin,C., de Francesco,M., Buell,G., Feger,G., Garcia,E., Peltsch,M. and Garcia-Bustos,J.F.
 TITLE Direct Submission
 JOURNAL Submitted (31-OCT-2000) Research Department, Glaxo Wellcome, S.A., Severo Ochoa 2, 28760 Tres Cantos, SPAIN
 COMMENT * NOTE: This is a 'working draft' sequence.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 FEATURES
 source
 1..11362
 /organism="Streptococcus pneumoniae"
 /serotype="19F"
 /db_xref="taxon:1313"
 /clone="G54"
 BASE COUNT 3304 a 2518 c 2123 g 3414 t 3 others
 ORIGIN
 Alignment Scores:
 Pred. No.: 13.4 Length: 11362
 Score: 133.00 Matches: 116
 Percent Similarity: 33.62% Conservative: 77
 Best Local Similarity: 20.21% Mismatches: 194
 Query Match: 3.58% Indels: 187
 DB: 2 Gaps: 26
 US-10-008-355-2 (1-712) x SPNEU1923 (1-11362)
 QY 131 LysIleValIleValIleThrAspLysValGluGlyGlnLeuLysGlyIleThrAspGluMet 150
 DB 5039 AAAATPAAATATACGATTTCCTCGAGAAAGACGTGCTAACCAATTCMAAATAATAAATA 4980
 QY 151 GluArgLeu-ArgLysAlaGlnGluValGlyGlnGlnLeuAlaLysLysGluAsnAlaAsn 170
 DB 4979 AACAAATATGAAAGACACACACATGCTCTAAGACTTTCATCAATTAATACAAATCCACG 4920
 QY 170 pEluAsnGlnLeuCySIIleValGluProPheTyrSerAsnAsnGluTyrPheLeuIleVal 190
 DB 4919 CGATGTTGAGGCTGCTGTACCAAAATG- 4884
 QY 190 IlyrAspValPhe- 4824
 DB 4883 TGCAGTCTTTCACAGCCCTTCAGCGCATCAAAAGCTTACCTTATTCATGCTGATTCC 4824
 QY 202 aProPheSerValGlyLysPheGlyLysPheAspAsnTyrPheMetIlePheArgIle 222
 DB 4823 ACCACCAACGTAACCTGGAACCTT- 4797
 QY 222 sThrGlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgProAlaGluTyr 242
 DB 4796 CCTTGT- 4788
 QY 242 rSerLysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGlnGlyTyr 262

DB 4787 CGCTGGGATACG---ACTTTCAGAGATATCATTCCTGTCMAAAACCGATCCAGCGCTT 4731
 QY 262 rLysAlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrLeu- 281
 DB 4730 TGAAT- 4689
 QY 282 -ThrSerTyrGlyValGluAspArgIleGluAsnGlnAsnAsnProArgIleGluValAr 301
 DB 4688 GACTCAGCGCTAGTATGAGAGACGCTTACGTGTGAGGGATTATACGGTATGACCTGAG 4629
 QY 301 gGly- 317
 DB 4628 TCGTAGTCTTCTTGACCAAGCTCTGGCAATGGAAGAC- 4581
 QY 317 aThrArgIleLysTyrAla--SerLysTyrAlaGlnSerAlaAsnTyrTyrLysAsnSe 336
 DB 4580 CACTACACATCAAGAAACAAATGGGCAAGATGGGGCTTCTGTAGACTATTCGCGAGCG 4521
 QY 336 rIleGlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGluGlu 356
 DB 4520 TTTCACTCTTGACGAAGGTTGTCA- 4485
 QY 356 ValArgAlaPheAlaAspTyrPheArgLys- 372
 DB 4484 TTAGGCTCTTGTGACCTTTACAAAGAGCTGATCATCTGCTGGATTTATCATCA 4425
 QY 372 rGlyAsp- 384
 DB 4424 CTGGGACCCAGACGCTCGACAGCCCTTCTGTATTTGAGGATTCACAAAGATGTAGA 4365
 QY 385 - 393
 DB 4364 AGCTGCTTCTACACATGATTTACATGCTGGAGATGCTTACCGCCCTTGAAGTTGC 4305
 QY 393 rThrTyrLeuSerGluThrLeuPheGlyLysThrGluValValArgPheAlaGlnPheAl 413
 DB 4304 TACAACTGCTCTGAGACTATGTTGGGACGTT- 4271
 QY 413 aAsnAlaLeuAlaTyrAsnProAspAlaHisAlaGlyIleLeuLysSerLeuAspAspLys 433
 DB 4270 -CGCGTGGATTAATCCAGAA- 4242
 QY 433 sTyrLysAsp- 447
 DB 4241 CTACAGAGCTGATGTTGTAATAACGTCATCTTCCA--ATCGCTAATAAACGTCATCCC 4185
 QY 447 CalMetLeuAsp- 454
 DB 4184 AATCGTTGGAGATGAGACGACGATCTGAGTTGGTACTGCTGGAATAATCACACC 4125
 QY 454 GArgArgIleProAlaAspLysLeuProAspIlePheLysAsnValIleAspLysLysP 474
 DB 4124 TGCCACAGCTCAATGACTCTTGTGGCCCAAGCTCATACTTGCACCAAGTCAACGT 4065
 QY 474 eLysGlyAspThrLysLysTyrAlaAspPheValPheAspLysSerValValProTyrSe 494
 DB 4064 CATGAACGACGACGAACATGATGAGCTTGTGAATTTTCAGGCATG- 4013
 QY 494 rAspLysPheHisAlaMetLeuLysSerMetAspLysGluLysPheAlaValIleGlu 514
 DB 4012 -GATGCTTTGAAAGCTCGT- 3969
 QY 514 uLysAspProAlaValIleGluSerLysSerValIleAlaAlaAlaVal- 531
 DB 3968 AATCGGTCCTTGTCAAAATGAAACGTTCCATTCAAGTCGTCACCTCAAGAACGAC 3909
 QY 532 - 533
 DB 3908 AGCTGCTGATGAGCCACGCTTGTACTCAATGATGTTGTCAGATGAGCAACATTTGCC 3849
 QY 533 nAlaAspAlaMetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPhePheAlaGlu 553

Db 3848 TAAGAACGCATTGCCAACACAGACAGAGACAAG----- 3812
Qy 553 ylenatgglumetfyrproglaralaleuproser----- 565
Db 3811 ---GTGGAATCTACCCACCTGTTCAAGCATCTCTCAATGATGAGAAATGT 3756
Qy 566 ---Aaplaasaphetfmetarwmetserfyriglsarilelysglytyrcluprogl 584
Db 3755 CCACGAGCTGGTTATCTCTGTCGACGTTGTGGGTGAC-----CA 3714
Qy 584 naspjglatatpyrasnrythisthrthrglysglyvalleuglulysglaspr 604
Db 3713 AATCCCTGCTGACCAAT-----CGTGTGTGAATGTATGTGCGGAGAACCTCC 3660
Qy 604 olysserasgluphealavalglnglulnleleuasp 617
Db 3659 AGAAGGTAGAGATGAGTCAGACGAGACGCTTGAT 3620

RESULT 9
LOCUS AE009858 14169 bp DNA linear BCT 16-JAN-2002
DEFINITION Pyrobaculum aerophilum strain IM2 section 113 of 201 of the complete genome.
ACCESSION AE009858 AE009441
VERSION AE009858.1 GI:18160615
KEYWORDS
SOURCE Pyrobaculum aerophilum.
ORGANISM Pyrobaculum aerophilum.
Archaea; Crenarchaeota; Thermoprotei; Thermoproteales; Thermoproteaceae; Pyrobaculum.
REFERENCE 1 (bases 1 to 14169)
Fitz-Gibbon,S.T., Ladner,H., Kim,U.J., Stetter,K.O., Simon,M.I. and Miller,J.H.
Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum aerophilum
Proc. Natl. Acad. Sci. U.S.A. 99 (2), 984-989 (2002)
PUBMED 11792869
JOURNAL 2 (bases 1 to 14169)
REFERENCE 2 (bases 1 to 14169)
Fitz-Gibbon,S.T., Ladner,H., Kim,U.-J., Stetter,K.O., Simon,M.I. and Miller,J.H.
Direct Submission
Submitted (12-DEC-2001) Microbiology and Molecular Genetics, University of California, Los Angeles, 405 Hilgard Ave, Los Angeles, CA 90095-1489, USA
JOURNAL
AUTHORS
TITLE
FEATURES
source
Location/Qualifiers
1..14169
/organism="Pyrobaculum aerophilum"
/strain="IM2"
/db_xref="taxon:13773"
144..1019
/gene="PAE2116"
144..1019
/gene="PAE2116"
/note="Protein fate; Protein modification and repair"
/codon_start=1
/transl_table=11
/product="methionine aminopeptidase"
/protein_id="AAL63956.1"
/db_xref="GI:18160616"
/translation="MLRLKQYGDVNVKALKYALDLTOPDMFVLELCERVEAFIRAND AKPAPVAVSVNEVAHYAKRQDSIKIPKGLVIVDGAQDGYIVDAAYVTIGPV FNNLOKAKSALESAINAKPGIKAMQIGEVVERITKSGLSPIYNLGHKVERYLH AGHVIPTYSDKASQALAGDYVAIEPFAITNGEYGVVDGGETTIRLARMHKNYOH IDIISAEAGLPDPFRMFPOLTDDAIRAKLAKGVAGYEIVLVERSKGYAPQEDTYVY GENEVIPLYRTIELL"
1060..2709
/gene="PAE2117"
1060..2709
/gene="PAE2117"
/note="Protein fate; Protein folding and stabilization"
/codon_start=1
/transl_table=11
/product="thermosome (chaperonin) alpha subunit"

gene
CDS
/protein_id="AAL63957.1"
/db_xref="GI:18160617"
/translation="MSQAVITQIGCVPLVLKEGTORAFKGAELRLNIMIRALAEV MRTLGPKMDKMLIDSLDITITNDGATILDEMDQHPITALLVEISKSQEEAGDGT TTAVALAGALIEAEKLEADKNHPVIVSGFRKALDVAEHLRVAIPVNTDYLTK RIAMTSMGKISEIETVEKEXFADLAVAVQVAEERNGKVDYDNIQIYKKGSLDR OLIVGIVVDKEVYHNAAMPKRVYNAKITALDLAEKEKEIDEIETINPTOMRALEE EERLRGVYDKLSIGVVLPTTKGIDIDIAQYILAKAGILAVRRVRSRSDIKRVANG ARLVTSIEDLTEADGFRAGLVEERVRGEKVFVQCKNPKAVSLIYGGCFERLYDEA ERNLDALSVSDVEDPYILPAGGAALIEAKAVARAPRVGGEQYAVEFAFALE VIPKALIAENAGLDPIDILTELFHKHEQDGMKYGIDVYQGRVMDVSLGLEPVLKI NALKVAVAASATLRIDEIIRASIKLEKEEKEEKEEKEEKEEKEEKEEKEEKEE complement(2706..3257)
/gene="PAE2118"
complement(2706..3257)
/gene="PAE2118"
/note="Hypothetical"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAL63958.1"
/db_xref="GI:18160618"
/translation="MLAEYVAORVALIGERAVYHRLVILSATSMSSVMMSGIALFEF AFKYLSIALALAGVLSRVLNMAEKKIRKPMANVEREMARRYPATLKEYPNDYL TKLVAEPTIDEPDIIAEEKKIRASTYLFPSQSITALRAKNRLVITITALLTISPHLVN KDIIAGVILILVILFRKRSVY"
complement(3247..4278)
/gene="PAE2119"
complement(3247..4278)
/gene="PAE2119"
/note="Unclassified"
/codon_start=1
/transl_table=11
/product="conjugal transfer protein, conjectural"
/protein_id="AAL63959.1"
/db_xref="GI:18160619"
/translation="MYPIIKRLLECAECCSSVEKIGCIDFAKEELPIYIKLSRYEKK SIDERLDFRGVLRIRNEKMAIRGLAVLAGEESLPLEDRVEDIVIVPGPIYTR KSGERKAVPAETKRVKSLIKAIYAKGVELTNSXGLKGLDRIASIDLPVY PVPQAVIRIHRKLVADMLKSGFTTEBQKLVAVMEGRHIVITGPGSGKITLLS VIDLIIPQLORVYIDEADDEDDPKNQIKIRNVKLEVYASLNRIIDIIIGELQ YEEHNFARKTAVEIGLQTLATMHATSVKDALRRLRGRISRENLAIIOLARKYDEIGE RRVVELYAS"
complement(4250..4729)
/gene="PAE2120"
complement(4250..4729)
/gene="PAE2120"
/note="Transcription; Transcription factors"
/codon_start=1
/transl_table=11
/product="transcription initiation factor IIB, conjectural"
/protein_id="AAL63960.1"
/db_xref="GI:18160620"
/translation="MKCPYQOSMTITLADGEVYCHDCTVLGPEVVPBPRLVLPVY RHKLIMLLEQENKSIKKKYSEWKMVTKGVAALGVPEVYVVALQIFORDRKIYO KSPRVIAAAYVALAERLGIYIHQVIAKILGVSKEFIRQASLRKRYVPRIKETT"
4857..5171
/gene="PAE2121"
4857..5171
/gene="PAE2121"
/note="Hypothetical"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAL63961.1"
/db_xref="GI:18160621"
/translation="MEFSLAVYDVTDNDMPNVYIKLGDVEATETPLEKIGVFSVGD R IKIVQKREKDDLNKXYKVMGIYVYVNGMTRISIGLQIDIMKELPKKGEKIYV IL"
5207..5641
/gene="PAE2122"
5207..5641

```

/gene="PAE2122"
/note="transcription: RNA processing"
/codon_start=1
/transl_table=11
/product="small nuclear ribonucleoprotein homolog
(sm-like)"
/protein_id="A163962.1"
/db_xref="GI:18160622"
/translacion="MASTLAASKRFEVLAELNLDLGRVQVLSNGEYKGVLAHVNDQ
LNIIVANASKNAGEKRFVIMRYIVHIDSTERRIDMREFAOAEKIFPGMKYIEE
TNVVLIDKRVSEIGVEGVPAAERAKRIFEEFLKRYSLQG"
complement(5871..6710)
/gene="PAE2125"
complement(5871..6710)
/note="Hypothetical; Conserved within genome"
/codon_start=1
/transl_table=11
/product="pAREP2a"
/protein_id="A163963.1"
/db_xref="GI:18160623"
/translacion="MTYKAVKACVVALPEAFSLPKKFGVSGCFTGCAFGVDAYI
AEGGRKIRFRVYLPDTIAKRYMAEKPKKRNREGICSTPLAELAYELAKNIEM
IEKRTKTYTYAVVDGVRLEGSICYSRESLGCISEVLRKVEQREWEQURORIRE
VLNALTIEWYRCFGETPAHPLKVAKRLEEFHALCOMSAKREFGIALPMERLRGAFW
WEGEMRKSMSCFTEKRAVCYKGEKMAAFYEDTPHGYLREPKLVDMIRIVAHNG
DSS"
complement(6700..7077)
/gene="PAE2127"
complement(6700..7077)
/note="Hypothetical; Conserved within genome"
/codon_start=1
/transl_table=11
/product="pAREP2a"
/protein_id="A163964.1"
/db_xref="GI:18160624"
/translacion="MAALMDYRCFKNPIMPEEGKLVRIELYYGCMEMATAIAE
YGEYKAPLISEVALRKAFWEGEMRKPMSCHTEKRAVCYKGMKATFYVDTPHG
VYLREPKLIDDMIKVAYRGDS"
7235..8263
/gene="PAE2128"
7235..8263
/note="Hypothetical; Conserved within genome"
/codon_start=1
/transl_table=11
/product="pAREP2b"
/protein_id="A163965.1"
/db_xref="GI:18160625"
/translacion="MEBAPGICIAVEEAVKPAVVAAGLAGYAALHDGLYSTAVSAT
AAAVLARREGAFERVAEYVRAAFAAEAREVEEKARITLORLEYELFVAVAVIDY
VRAHWFTMAAAGIIMAVVAAQLODYLIMNHIALNAGALAGAKAGVCDKRGVN
AVUTKATSEKEIYBRAAOLGIDGQVYKKAISASTLRHNDKQIEKAVYELKRLSY
VQAKKAAHYKALEEASKAMYEDDPALALALVGLARETGKSVKFEEDDDDIATYIT
ASFTLNHLREFELKNKSYKALERLEGAVKSGRFALSDVWGLNLVEAAERAKGY
ORAGQONCI"
8307..11282
/gene="PAE2129"
8307..11282
/note="Hypothetical; Conserved within genome"
/codon_start=1
/transl_table=11
/product="pAREP2b"
/protein_id="A163966.1"
/db_xref="GI:18160626"
/translacion="MATENVIRLAEARDDLPKIDATRGERTVATLLSLVTGLFGVT
VERALARGKMTVGALGASAVRSTPSNFNASSVSGDEKPSAKKLTFRIRIATLAD
PAMEIILAGQGVVEQKTEKNGKTYITFDKDHLLKAPMDRGLMPMAAG
EVELFEVGLKAVVSSGSKPLEELNEEMKRVVEVRVKKAIEMTVAAVIGALP
TDVAILPGSEYVKGTSYLSQTFYTMALADRGLVAVYSEBGLKFMKMEVEKTYTE
VREVLVSRAVFENLTPKROSGEVDLKVALTVDVKNDELKALKEAFMEWERVE

```

```

Alignment Scores:
Pred. No.: 22.5 Length: 14169
Score: 131.50 Matches: 138
Percent Similarity: 34.63% Conservative: 112
Best Local Similarity: 19.11% Mismatches: 247
Query Match: 3.54% Indels: 225
DB: 1 Gaps: 36

US-10-008-355-2 (1-712) x AE009858 (1-14169)

QY 111 ValSerArgThrMetGlyGluGluLeuProIlePro-----GlyLeu 124
|||||
DB 9153 GTAAGCCGCCGCCGCTTTGAATAATCTATGACACCAACAAGCTGGAAGCGCGCTT 9212

QY 125 SerValLysTyrLeuArgLysIleValLysValThrAspLysValGluGlnLeuLys 144
|||||
DB 9213 GATCTGAAGGTGGCTCTACTGATGATTAAATGATGATGAATTAAGAAAGCGCTTGAG 9272

QY 145 GlyIleThrAspGluMet-----GluArgLeuArgLysAlaGlnGluValCysGlnGluLeu 163
|||||
DB 9273 ACGGCGCGCGCGCGAGTCTGCGAGCGCGCTTATAGCTGTTAGCACTGTGGAACACTGA 9332

QY 164 AlaLys-----LysGluAsnAlaAspGluAsnGlnLeuCysIleValGluProPheTyr 181
|||||
DB 9333 GAAGAGAGAGCAAGAGCAAGAAAGAGCAAGAGCAAGAAAGGTTAATGAGAGTAGAC 9392

QY 182 SerAsnAsnGluTyr----- 186
|||||
DB 9393 AAGCTGGCAAAATACCTCCGCGTCTCTCCGCTGCATATGCGGTAGAGCGCGCAAG 9452

QY 187 -----PheLeuIleValTyrAsp 192
|||||
DB 9453 ATGGGTGAAGTCTCTAGAGAGAGAGCGCGCGCTTGGCGTGAATTTGCGTGTTTACGAT 9512

QY 193 -----ValPheLysAspValArgMetValPhe-----Ala 202
|||||
DB 9513 GCGACTGTGTTAGAGCAAGAAATCTCTCTTTCATCCGCGCGCGCTGAACAGTGAAGAG 9572

QY 203 ProProSerSerValGlyLysPheGlyGlyAspThrAspAsnThrMetTyrProArgHis 222
|||||
DB 9573 CCGATTATGACCGCGCACCACTTC-----ACCGCTTCTGCGTGTGGCTCTGAG 9623

QY 223 ThrLysAspPheSerValPheArgValTyrAlaGlyLysAspAsnArgProAlaGluTyr 242
|||||
DB 9624 GAGCTTGCGCTTAAGCAAGCTGATGATGAGGC-----AAGGGCGGCACCAT 9674

QY 243 SerLysAspAsnLysProTyrLysProValTyrPheAlaAlaLysMetGlnGlyTyr 262
|||||
DB 9675 -----ATTGCCCTTCAAGCGCGCGCGCTGACGAGCTTAGGGATTTA 9707

QY 263 LysAlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAsp----- 278
|||||
DB 9708 CTGAG-----GCTGTACGCGCGCGCTACCGCGCTTGACGAGCTTAGGGATGCC 9758

QY 279 -----ArgTyr 280
|||||
DB 9759 CTGACAGATTCGCGCAGCGCTTAAAGTCGTACCGCGCAGCGGCTTAAAGCAAAATTC 9818

QY 281 LeuThrSerTyrGlyValGluAspArgIleGluAsnGluAsnAsnProArgIleGluVal 300
|||||
DB 9819 GGCATTGACCTGGCTTACGACGTGAGAAACGAATGTTTTCAAAAAGCTAGAGGAAGTT 9878

QY 301 ArgGlyIleLysGlnGly-----IleTyrLysGlu-----AlaMetSerAlaAspGln----- 316
|||||
DB 9879 GTTCAATGGCAGAGACTACGCTACAGAAACCTCACAGTGGAGAGGATCACTGAGAC 9938

QY 317 -----AlaThrArgIleLysTyr-----AlaSerLysTyrAla 327
|||||
DB 9939 ACAAGCGGACAGCTTGCAAGAGCTGTCATCCGCTTAAAGCTTGCGGGGAGAGAGTGGCG 9998

QY 328 GlnSerAlaAsnTyrTyrLysAsnSerIleGlyMetAsnArgGlyLeuAlaArgLeuAsp 347

```

```

Db 9999 TACATAACCGTATACGTGACGTGCACT-----AGCTACTAGGCCAATTCGAC 10046
Qy 348 ValIleGlyArgLysArgAlaGluAlaAlaPhe----- 359
Db 10047 GGC---TCTCGAAGAAAAAGCGCAAGTATGCTTATCATTCAGCGCTTTGGCGGCAAG 10103
Qy 360 AlaSPTrpIleArgLysAsnGlyLysSerAlaValTyrGlyAspValLeuSerSerLeu 379
Db 10104 GCGGAGGTCAACCAAGATGCGCAATGACCTGTTAC----- 10142
Qy 380 GluLysAlaTyrLysGluGlyAlaLysAlaAsnArgLysMetThrTyrLeuSer----- 397
Db 10143 -----CTCCACACCGCGCATTAATTTGCCATTCGCCATGCAGCGTGAATATGCTGTG 10196
Qy 398 -----GluThrLeuPhe-----GlyGlyThrGluValAlaArgPheAlaGln 411
Db 10197 AAGCGTTTGTTCACCAACTTTCAGTAAAGGGCTGATCAGCAGAGATAGATACGAACAG 10256
Qy 412 PheAlaAsnAlaLeuAlaThrAsnProAspAla-----HisAlaGlyIleLeuLysSer 429
Db 10257 TTGCTGAGGACATAGCTACAGGCCCAACACCGCTCACTATGCGGGGTGGAGTTCTCG 10316
Qy 430 LeuAspAspLysTyrLysAspTyrLeuProSerLeuAspArgLysValLeuProAlaMet 449
Db 10317 GTTTATCACAAAGCAGAGAGTAATAGATTTTATGAATAAGTATACCCAGGC--- 10373
Qy 450 LeuAspIleValArgArgArgIleProAlaAspLysLeuProAspIlePheLysAsnVal 469
Db 10374 -----ANCGAGCGCTCTCAAAAATGCG 10394
Qy 470 IleAspLysLysPheLysGly-----AspThrLysLys 480
Db 10395 GTTGTGAACGCCCTAAAGCAAGAGGCTGAAGAGGCTACACTTCACTGTGAAGAG 10454
Qy 481 TyrAlaAspPheValPheAspLysSerValValProTyrSerAspLysPheHisAlaMet 500
Db 10455 TACGGCGACACTACGAATTCACGTGCTGTGAATCGATACAAAGCCTTGGAGACCTTG 10514
Qy 501 LeuLysSerMetAspLysGlyLysPheAlaLysAlaIleGluLysAspProAlaValGln 520
Db 10515 GCACGATGCTGGCTGAGAAAGGAGAGTACTACGCCATTCAT----- 10556
Qy 521 LeuSerLysSerValIleAlaAlaAlaArgAlaIleGlnAlaAspAlaMetAlaAsnAla 540
Db 10557 ----GACAGAAAGCGCTGAATACAGCTCAAGCGGAGCAGCAAAAGCAGTGTGAATACC 10613
Qy 541 TyrAlaIleGluLysGlyLysArgLeuPheAlaGlyLeuArgLysMetTyrProGly 560
Db 10614 -----CTAAAAACCGCTGGCTGGAGGAG----- 10637
Qy 561 ArgAlaLeuProSerAspAlaAsnPheThrMetArgMetSerTyrGlySerIleLysGly 580
Db 10638 -----GATACACACTTTCACGTGAGTAGTG----- 10661
Qy 581 TyrGluProGlnAspGlyAlaTTrpTyrAsnTyrHisThrThrGlyLysGlyValLeuGln 600
Db 10662 -----GACGGA---TACTACGTAAATTCGTAATACATAGAAAGCCTCCGCGAA 10706
Qy 601 LysGln-----AspProLysSerAspGluPheAlaValGlnGluAsn 614
Db 10707 ATCCAGAGTATGGCGCTAAACGGCGACGACGTGAGCGCTGATTATA-----AGGAG 10760
Qy 615 IleLeuAspLeuPheArgThrTyrLysAsnTyrGlyArgTyrAlaGluAsnGlyGlnLeuHis 634
Db 10761 CTTGAGGATGTGCTGAGG---CGCCCGCACGGCGCAAAACGACAGTGAAGAGCGATTTGAA 10817
Qy 635 IleAlaPheLeuSerAsnAsnAspIleThrGlyLysAsnSerGlySerPro-----Val 652
Db 10818 GTGTTCAGCGCGCGAGGAGGAG-----GGACGATAGATTTTACCGCTAGAAAGTG 10868
Qy 653 PheAspLysAsnGlyArgLeuIleGlyLeuAlaPheAspLysAsnTrpGluAlaMetSer 672

```

```

Db 10869 CGCGACGAAAGGGGCAACGCGTAGCCGCGTGTGAGCTTGAGTATGATTCGTAAATG 10928
Qy 673 GlyAsp-----IleGlu 676
Db 10929 GCGGATCAACCCGTTGGCAGTGCAGCGCGGAGAGACTGCCCTGGCATTAATAGGAG 10988
Qy 677 PheGluProAspLeuGlnArgThrIleSerValAspIleArgTyrValLeuPheMetIle 696
Db 10989 TACGACACAGAGGGGGAAGAGAG-----CACCTAAAGATAGTATGTTACTGCGTT 11039
Qy 697 AspLys 698
Db 11040 AAGAAA 11045
RESULT 10
A81135
LOCUS A81135 2652 bp DNA linear PART 21-JAN-2000
DEFINITION Sequence 1 from Patent EP0909818.
ACCESSION A81135
VERSION A81135.1 GI:6731609
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 2652)
AUTHORS Brown, J.R. and Wang, M.
TITLE Valyl CRNA synthetase (EC 6.1.1.9) from Streptococcus pneumoniae
JOURNAL Patent: EP 0909818-A 1 21-APR-1999;
SMITHKLINE BEECHAM CORP (US)
FEATURES
source 1..2652
location/Qualifiers
BASE COUNT 733 a 612 c 621 g 686 t
ORIGIN
Alignment Scores:
Pred. No.: 3.68 Length: 2652
Score: 130.00 Matches: 106
Percent Similarity: 32.818 Conservative: 62
Best Local Similarity: 20.708 Mismatches: 166
Query Match: 3.508 Indels: 178
Gaps: 25
US-10-008-355-2 (1-712) x A81135 (1-2652)
Qy 192 AspValPhe-----LysAspValArgMetValPheAlaPro 203
Db 76 GATGTTTTCAGGCTTCAGGCGATCAAAAGCCTAAGCCTTATTCATTCGTTATTCACCA 135
Qy 204 ProSerSerValGlyLysPheGlyGlyAspThrAspAsnTrpMetTrpProAlaGlnHisThr 223
Db 136 CCAAAAGTTACAGTAAACTT-----CACCTT 162
Qy 224 GlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgProAlaGluTyrSer 243
Db 163 GGT-----CACGCT 171
Qy 244 LysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGlnGlyTyrLys 263
Db 172 TGGGATACA---ACTTTCGAAGATATTATCATCCGTCGAATAAGCATGCAAGTGTGAT 228
Qy 264 AlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrLeu---Thr 282
Db 229 -----ACCCCTTGGCTTCCCTGGGATGACCCACGACGAGGATTCGCACT 270
Qy 283 SerTrpGlyValGluSparGlyIleGluAsnGluAsnAsnProArgIleGluValArgGly 302
Db 271 CAGGCTAAGGTAGAGAGCGCTGCGTGTGAGGCAATTTCCGCTATGACCTTGGTCTG 330
Qy 303 -----IleLysGlnGlyIleTyrLysGluAlaMetSerAlaAspGlnAlaThr 318
Db 331 GAGCTTTTCTTGACGAAAGCTCGGAATGGAAGAC-----GAATATGCACT 378

```

```
Oy 319 ArgIleIysTYrAla---SerIysTYrAlaInserAlaasnTYrIlePlysasnSerIle 337
      |||||
Db 379 ACTATCAAGAACATGGGGCAAGATGGGGCTCTCTGTAGACTATTTCTGTAGCGTTTC 438
Oy 338 GlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgIysArgAlaGluIuArg 357
      ::::: |||||:::
Db 439 ACTCTTGACGAAGTTTGCA-----AAAGCTGTCGTAG 474
Oy 358 AlaPheAlaAspTPrIleArgIlys-----AsnGlyIysSerAlaValIleArgIly 373
      ||| ||| ::::|
Db 475 GTCCTTGAGACCTTTACAGAAAGCGTGATCTACCGGTGGAGTTATATCTCACTAG 534
Oy 374 Asp-----ValIysSerIleGluIysAlaTYrIysGluIuIysAlaIys 388
      ||| ||| ::::|
Db 535 GACCCAGCAGCTCGACAGCCCTTTCTGTATATTGAGTGATTCACAAAGATGTGAGAGT 594
Oy 389 AlaAsnArgGluMetThrTYrLeu----- 396
      ||| ||| ::::|
Db 595 GCCTTCTACACATGATTAATTCATGCTGGAAGATGGTTCACGGCTCTGAAGTGTCTACA 654
Oy 397 -----SerGluThrLeuPheGlyGlyThrGluValAlaArgPheAlaGlnPheAlaAsn 414
      |||||:::|||||
Db 655 ACTGCTCTGAGAGCACATGTTTGGGACGTT----- 684
Oy 415 AlaLeuAlaThrAsnProAspAlaHisAlaGlyIleLeuIysSerLeuAspIysTYr 434
      |||||:::|||||
Db 685 GCGGTTGGCGCTCAACCCAGAA-----GACCCGGCGTAC 717
Oy 435 LysAsp-----TYrLeuProSerLeuAspArgIysValIleuProAla 448
      |||||
Db 718 AAGGACTTGATGTGTAATAATGTCATCCCTTCCA---ATCGCTAATAAATCTCCCAATC 774
Oy 449 MetLeuAsp-----IleValAlaArgArg 455
      ::| |||
Db 775 GTTGGAGATGACAGCAGATCTCTAGTTTGCTACTGGTCTGTAANATCCACACCTGCC 834
Oy 456 ArgIleProAlaAspIysLeuProAspIlePheIysAsnValIleAspIysIysPheIys 475
      ||| ||| |||
Db 835 CACGATCCAAATGACTTCTGTGGTGGCCAAAGTCATTAAGTCCCAAGATCAAGCTCATG 894
Oy 476 GlyAspThrIysIysTYrAlaAspPheValIleAspIysSerValIleValProIysSerAsp 495
      ||| |||
Db 895 AAGCAGCAGCAGAACATGATGATGCTGGCTTGAATTTTCAGGCAAG-----GAC 945
Oy 496 LysPheHisAlaMetLeuIysSerMetAspIysGluIysPheAlaIysAlaIleGluIys 515
      ::||| ||| |||
Db 946 CGTTTGAAGCTCGT-----AAGCAGCTGCTTGAAGTGTGAAGAAATC 990
Oy 516 AspProAlaIleGluLeuSerIysSerValIleAlaAlaAlaArgAla----- 531
      |||||:::|||||
Db 991 GGTGCCCTGTCMAAATGCAAAACGTGTCACAGTGTGTCACCTCAGAGCGTACAGGT 1050
Oy 532 -----IleGlnAla 534
      ::|
Db 1051 GTTGGTGTGAACCTGCTGTCTACTCAATGCTTGTCAAGATGACCAATGGTCTAG 1110
Oy 535 AspAlaMetAlaAsnAlaTYrAlaIleGluIysGlyIysArgLeuPhePheAlaGlyLeu 554
      ::|||:::|||||
Db 1111 AAGCGCATTTGCCAACAGACACAGACAGCAAG----- 1143
Oy 555 ArgGluMetTYrProGlyArgAlaLeuProSer----- 565
      ||| ||||| |||
Db 1144 GTCGAATTCACCCACACCTGTTTCAACGATACCTTCAATGATGGAATAATGTCCAC 1203
Oy 566 AspAlaAsnPheThrMetArgMetSerTYrGlySerIleIysGlyTYrGluProIleAsp 585
      ||| ::|
Db 1204 GACTGGGTTATCTCTGCTGAGCTGCGGGGTGAC-----CAAATC 1245
Oy 586 GlyAlaTYrTYrAsnTYrHisThrGlyIysGlyValIleGluIysGlnAspProIys 605
      ||||| |||||
Db 1246 CCTGCTGTGTACAAAT-----GCTGATGTGTAAATGTATGTGCGGCAAGAACCTCCAGAA 1299
```

```
Oy 606 SerAspGluPheAlaValAlaGluIuAsnIleLeuAsp 617
      ||| ::|
Db 1300 GGTGACGATGTGACTCAGCAGCAAGACGCTTGTGAT 1335
RESULT 11
LOCUS AR065782 2652 bp DNA linear PAT 29-SEP-1999
DEFINITION Sequence 1 from patent US 5849555.
ACCESSION AR065782
VERSION AR065782.1 GI:5959598
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 2652)
AUTHORS Brown,J.R., Jaworski,D.D., Lawlor,E.J. and Wang,M.
TITLE Valyl Lrna polynucleotides from Streptococcus pneumoniae
JOURNAL Patent: US 5849555-A 1 15-DEC-1998;
FEATURES
source 1..2652
BASE COUNT 733 a 612 c 621 g 686 t
ORIGIN
Alignment Scores:
Pred. No.: 3.68 Length: 2652
Score: 130.00 Matches: 106
Percent Similarity: 32.81% Conservative: 62
Best Local Similarity: 20.70% Mismatches: 166
Query Match: 3.50% Indels: 178
DB: Gaps: 25
US-10-008-355-2 (1-712) x AR065782 (1-2652)
Oy 192 AspValPhe-----LysAspValArgMetValPheAlaPro 203
      |||||
Db 76 GATGTTTCAAGCCTTCAGCGGATCAAAAGCGTAAAGCTTATTCATGCTTATTCACCA 135
Oy 204 ProSerSerValGlyIysPheGlyGlyAspThrAspAsnTYrMetTrpProArgHisThr 223
      |||||
Db 136 CCAACGTTACAGTAACT-----CACCTT 162
Oy 224 GlyAspPheSerValPheArgValTYrAlaGlyAlaAspAsnArgProAlaGluIysSer 243
      |||
Db 163 GGT-----CACGCT 171
Oy 244 LysAspAsnIysProTYrIysProValTYrPheAlaAlaIleSerMetGlnIlyTYrIys 263
      |||
Db 172 TGGGATACA---ACTTGCAGAAATATATATCATCCGTCAAAAACGATGCAAGGATTTTGAT 228
Oy 264 AlaAspAspTYrAlaMetThrIleGlyPheProGlySerThrAspArgTYrLeu---Thr 282
      |||||:::|||||
Db 229 -----ACCCTTGGCTTCTCTGGATGGACACGACGAGGATGCCACT 270
Oy 283 SerTrpGlyValGluAspArgIleGluAsnGluAsnAsnProArgIleGluValArgGly 302
      |||||:::|||||
Db 271 CAGGCTAAGTAGAAGAGCGCTGCGTGTGAGGCATTTCCCGCTATGACCTGGTGTGT 330
Oy 303 -----IleIysGlnGlyIleTrpIysGluAlaMetSerAlaAspGluAlaThr 318
      ||| |||||
Db 331 GAGTCTTCTTGACGAAAGCTGCGGAATGAAAGC-----GAATATGCCACT 378
Oy 319 ArgIleIysTYrAla---SerIysTYrAlaInserAlaasnTYrIlePlysasnSerIle 337
      |||||
Db 379 ACTATCAAGAACATGGGGCAAGATGGGGCTCTCTGTAGACTATTTCTGTAGCGTTTC 438
Oy 338 GlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgIysArgAlaGluIuArg 357
      ::::: |||||:::
Db 439 ACTCTTGACGAAGTTTGCA-----AAAGCTGTCGTAG 474
Oy 358 AlaPheAlaAspTPrIleArgIlys-----AsnGlyIysSerAlaValIleArgIly 373
      ||| ||| ::::|
Db 475 GTCCTTGAGACCTTTACAGAAAGCGTGATCTACCGGTGGAGTTATATCTCACTAG 534
```


338 GlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGluGluArg 357
Db 439 ACTCTTGAAGAGGTTTCTCA-----AAGCTTCTGTAAG 474
338 AlaPheAlaAspTrpIleArgLys-----AsnGlyLysSerAlaValTrpGly 373
Db 475 GTCCTTGTGACCTTTACAGAAAGCGTGAATCTACCGTGGAGTATTCATCAACTCG 534
374 Asp-----ValLeuSerLeuGluGlyLysAlaTrpLysGluGlyAlaLys 388
Db 535 GACCCAGACAGCTCCGACAGCCCTTCTCATATGAGTGATTCACAGAGATGTGAAGCT 594
389 AlaAsnArgGluMetTrpTrpLeu----- 396
Db 595 GCCTTCTACACATGATTAATCATGCTGAGAGATGTTCCAGCGCTTGAAGTCTACA 654
397 -----SerGluThrLeuPheGlyGlyThrGluValValArgPheAlaGlnPheAlaAsn 414
Db 655 ACTCGTCTGTAGACCATGTTGGGAGCGTT----- 684
415 AlaLeuAlaTrpAsnProAspAlaHisAlaGlyIleLeuLysSerLeuAspAspLysTrp 434
Db 685 GCGGTGGGGGTCAACCCAGAA-----GACCCCGGCTAC 717
435 LysAsp-----TyrLeuProSerLeuAspArgLysValLeuProAla 448
Db 718 AAGGACTGATGTGTAATAATGTCACTCTTCCA--ATCGCTAATTAACCTCATCCCAATC 774
449 MetLeuAsp-----IleValArgArg 455
Db 775 GTTGGAGATGAGACGACGAGATCCGATGTTGGTACTGCTGTCGTAATAATCACACCTGCC 834
456 ArgIleProAlaAspLysLeuProAspIlePheLysAsnValIleAspLysPheLys 475
Db 835 CACGATCCAAATGATCTTGTGGTGGCCAGCTCACTTAAGTGGCCAGTCAAGCTCATG 894
476 GlyAspThrLysLysTyrAlaAspPheValPheAspLysSerValValProTyrSerAsp 495
Db 895 AAGCGACGACGAAACCATGATGACTGGCTTTGAATTTCCAGGCATG-----GAC 945
496 LysPheHisAlaMetLeuLysSerMetLysPheLysGluLysPheAlaLysAlaIleGluLys 515
Db 946 CTTTGAAGCTCGT-----AAGCGACGTCGTTGCTTAAGTTGGAAGAAATC 990
516 AspProAlaValGluLeuSerLysSerValIleAlaAlaAlaArgAla----- 531
Db 991 GGTGCCCTGCTCAAAATGCAAAAGCTGCCACAGTGTGGTCACTCAGAGCGTACAGGT 1050
532 -----IleGlnAla 534
1051 GTTGTGTTGAACCTCGCTTGTCTACTCAATGCTTCGTCAGAGATGACCAATTGGCTAAG 1110
535 AspAlaMetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPhePheAlaGlyLeu 554
Db 1111 AAGCCCATTTGCCAACCAAGACACAGAGAGCAAG----- 1143
555 ArgGluMetLysProGlyArgAlaLeuProSer----- 565
Db 1144 GTGCAATTCCTACCCACGCTTTCACAGATACCTTCCATCATGATGGAATGTCAC 1203
566 AspAlaAsnPheThrMetArgMetSerLysGlySerIleLysGlyTyrGluProGlnAsp 585
Db 1204 GACTGGTATTCCTGTCGACGCTGTGGGGGTAC-----CAATTC 1245
586 GlyAlaIlePtyrAsnTyrHisThrThrGlyLysGlyValLeuGluLysGlnAspProLys 605
Db 1246 CCTGCTGCTGATCAAT-----GCTGATGCTGAATATGATGCGGAGAGCTCCAGAA 1299
606 SerAspGluPheAlaValGlnGluAsnIleLeuAsp 617
Db 1300 GTTGACGATGACTCAGAGACGACGCTTTGGAT 1335
RESULT 13

CLOCOLOA
LOCUS CLOCOLOA 4141 bp DNA linear BCT 01-FEB-2000
DEFINITION C. perfringens colA gene for collagenase, complete cds.
ACCESSION D13791
VERSION D13791.1 GI:440850
KEYWORDS collagenase.
SOURCE Clostridium perfringens (strain:NCIB 10662) DNA.
ORGANISM Clostridium perfringens
Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
Clostridium.
REFERENCE 1 (bases 1 to 4141)
AUTHORS Matsushita,O., Yoshihara,K., Katayama,S., Minami,J. and Okabe,A.
TITLE Purification and characterization of Clostridium perfringens
120-kilodalton collagenase and nucleotide sequence of the
corresponding gene
JOURNAL J. Bacteriol. 176 (1), 149-156 (1994)
MEDLINE 94110220
REFERENCE 2 (bases 1 to 4141)
AUTHORS Matsushita,O.
TITLE Direct Submision
JOURNAL Submitted (28-NOV-1992) Osamu Matsushita, Kagawa Medical School,
Department of Microbiology: 1750-1 Ikenobe, Miki-cho, Kita-gun,
Kagawa 761-07, Japan (E-mail:omatsushedd@nig.ac.jp,
Tel:0878-98-5111(ex.2512), Fax:0878-98-7109)
COMMENT Submitted (28-NOV-1992) to DDBJ:
Osamu Matsushita
Department of Microbiology
Kagawa Medical School
1750 Ikenobe, Miki-cho
Kita-gun, Kagawa 761-07
Japan
Phone: 0878-98-5111 x2512
E-mail: omatsushedd@nig.ac.jp
Fax: 0878-98-7109.
FEATURES
source
1..4141
/organism="Clostridium perfringens"
/strain="NCIB 10662"
/db_xref="taxon:1502"
371..3685
/gene="colA"
371..3685
/gene="colA"
/codon_start=1
/transl_table=1
/product="collagenase"
/protein_id="BA02941.1"
/db_xref="GI:440851"
/translation="MKKNLKRGEITLKLVERSAFTTLAFTLFNSSFFVPAADKY
ENSNNGQITREINADQISTKELNNEVATDNNRPLGPSIAPSRANNKIYFEDLNRMN
YSDIVELIKTISEYNDPDLFPNDGSTFESNDRQAIYIGLEDSGRITVADDDGI
PTLVEFLRAGYVGEFYKQVLYNTPLKKECPAMKAIQYNSNFRIGTAAQGVVEA
LGRIGNASADPEVINCIVLVDFKQNDIKVGSNTSKGAVENTLKKGIDYVNSVY
NTGYDAKNTPEFNRIDPYMERLESCTTIDKLNNDNMLVNNALYYTGMGFRDP
SISORALERAMKEPYLSQYIEANDLDLNFSGKNSGNDIPNLIKADARKYLK
TYTFEDDKFVYKAGDKYTEKIRKLYWASKVAQEVAVVQNDKALEGNPDILTVY
IYNSPEBYKLNRIINGRSTNGCIYIENIGTFEYETPEESITLEELRHFEFTHL
OGRVYVGMGQGEFOEGVLTVWEEGTAEFFAGSTGDIKPKRSYTOGLAVDRNR
MSLYGVLAHYGSMDFYNGFALSNWYNNMGMFNMTVYIKNNVSGYKQDIKAS
SDYGLNDKDYDMSLNNIDNLDVPLVSEYVNGHEAKOINETDIKESVNTKIDS
SNVEKSFQFTTYMRGTYVGRSGOGEENDKDNKSLNDMLKELSKSNMGYFTVAY
FVNHKYDENGNYVYDVFGHGMNDTNTDVHNKEPAVIKSDSSVIVEEINFDTGS
KDEDEGELKAYEMDEGDEKSNKATPKYKTEGEYKLVLTQNNNGINTESSKIKY
EDKPEVYINSEPNNDPEKANOIAKSNMLYKGLTSEDDYKDYFPVAKGNKRTLN
NLSVGTITWTLTKEGDLNNTVLAITGDKTELKGEITLEPGRITLSTYTTIDNOSGAT
VNVKGLNKINVKETEKDAIKVENNNDPEKAMVVDNSKIVGTLSDDLIDYISDIO
NPDDLNTVVENLDNIKNNMLLYSADDLISNVVDVANADGNKLSNTCKLNPKGYLVAYVO
FENSGTGNVTYVNLQNK"
629..3682
/gene="colA"
/product="collagenase"
BASE COUNT 1672 a 426 c 735 g 1308 t
ORIGIN


```

QY 571 MetargmetSerTyrglySerIleuylSerIleuyluprogInaspGlyAlaTrpTyraSn 590
Db 3230 CCTGGAAGACTACTACTAGTATATATCTATATGATATCAATCAGAGCT----- 3280
QY 551 TyrHisThrThnGlyLysGlyValleuGluLysGlnaspProLysSeraspGluPheAla 610
Db 3281 TACACAGTAATGTATAAGCAACCTTAATAAGCTTAAGCAACCAAGCAAGCATCT 3340
QY 611 ValGlnGluAsnIleleuAspLeuPheArgThrLysAsnTyrglyArgTyraGluAsn 630
Db 3341 ATAAAGAA-----GTTGAATAAT 3358
QY 631 GlyIleuHisIleAlaPheLeuSerAsnAspIleThrGlyLysAsnSerGlySer 650
Db 3359 -----AACATATGAT----- 3367
QY 651 ProValPheAspLys-----AsnGlyArgLeuIleGly-----LeuAla 663
Db 3368 -----TTTGATAAAGCATATGAAAGGTAGACAGTAATAGCAAAATGTGTGAACATTAAGC 3421
QY 664 PheAspGlyAsnTrpGluAlaMetSerGlyAspIleGluPheGluProAspLeuAlaArg 683
Db 3422 AATGATGATCTTAAGGATATTATATGATATATCAAAATCCAAAGTCACTTAACATA 3481
QY 684 ThrIle-----SerValAspIleArgTyraValleuPheMetIleAsp 697
Db 3482 GTAGTTGAACCTAGATATATATAAATAAGAACTGTTATTATATTCAGCTGAT 3535

RESULT 14
BD003707/c 6004 bp DNA linear PAT 31-JAN-2002
DEFINITION Polynucleotide of Streptococcus pneumoniae and sequence.
ACCESSION BD003707
VERSION JP 2001501833-A 27 13-FEB-2001;
KEYWORDS JP 2001501833-A/27.
SOURCE unclassified.
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 6004)
AUTHORS Kunsch, C.A., Choi, G.H., Dillon, P.J., Rosen, C.A., Bars, S.C.,
Fannon, M. and Dougherty, B.A.
TITLE Polynucleotide of Streptococcus pneumoniae and sequence
JOURNAL Patent: JP 2001501833-A 27 13-FEB-2001;
HUMAN GENOME SCIENCES INC
COMMENT OS Unidentified
PN JP 2001501833-A/27
PD 13-FEB-2001
PE 30-OCT-1997 JP 1998520718
PR 31-OCT-1996 US 60/029960
PI CHARLES A KUNSCH, GIL H CHOI, PATRICK J DILLON, CRAIG A ROSEN, PI
STEVEN C BARASH,
PI MICHAEL FANNON, BRIAN A DOUGHERTY
PC C12N15/09, A01K67/027, C07K14/315, C07K16/12, C12N1/15, C12N1/19,
PC C12N1/21,
PC C12N5/10, C12P21/02, C12Q1/68, G06F17/30, C12N15/00, C12N5/00, PC
G06F15/40
CC Strandedness: Double;
CC Topology: Linear;
FH key Location/Qualifiers
FT source L. 6004
FEATURES
source location/Qualifiers
1..6004
/organism="unclassified"
/db_xref="taxon:32644"
BASE COUNT 1733 a 1296 c 1115 g 1860 t
ORIGIN

```

Alignment Scores: 11.8 Length: 6004
 Pred. No.: 129.00 Matches: 116
 Score: 33.62% Conservative: 77
 Percent Similarity: 33.62%

```

Best Local Similarity: 20.21% Mismatches: 194
Query Match: 3.47% Indels: 187
Db: 6 Gaps: 26

US-10-008-355-2 (1-712) x BD003707 (1-6004)
QY 131 LysIleValLysValThrAspLysValGluGlyIleuLysGlyIleThrAspLumet 150
Db 4315 AAAATTAATTAATACGAATTTCTGCAAGAACAGACTGCTTAAGCAAAATTCAGAAATAATA 4256
QY 151 GluArgLeu-ArgLysAlaGlnGluValCysGlnGluLeuAlaLysLysGluAsnAlaAs 170
Db 4255 AACAAATAGAGAAAGACACACACATCTTAAGAACTTCACTTAATCAATCCAGC 4196
QY 170 PGIuAsnGlnLeuLysIleValGluProPheTySerAsnAsnGluLysIleuLeu 190
Db 4195 CGAGTTGAGGCTGCTGCTTACCAAAATG-----CTTGA 4160
QY 190 LTyraPValPhe-----LysAspValArgMetValPheAl 202
Db 4159 TCGGATGTTTTCAGACCTTCAGCGATCAAAAGCTTAAGCTTATTCATCGATTC 4100
QY 202 aProProSerSerValGlyLysPheGlyGlyAspThrAspAsnTrpMetTrpProArgH 222
Db 4099 ACCACCAACCTAATCTGGAAACTT-----CA 4073
QY 222 sThrGlyAspPheSerValPheArgValTyraGlyAlaAspAsnArgProAlaGluTy 242
Db 4072 CCTTGGT-----CA 4064
QY 242 rSerLysAspAsnLysProTyrgLysProValTyraPheAlaAlaValSerMetGlnGlyTy 262
Db 4063 CGCTGGGATACG---ACTTTCAGAGATATCATCATCGTCMAAAGCAAGCATCAAGCTT 4007
QY 262 rLysAlaAspAspTyraAlaMetThrIleGlyPheProGlySerThrAspArgTyrl 281
Db 4006 TGAT-----ACCTTGGCTTCCAGATGTGACCAACCGCTGTATTCG 3965
QY 282 -ThrSerTrpGlyValGluAspArgIleGluAsnGluAsnProArgIleGluValAr 301
Db 3964 GACTCAGGCTAAGGTATGAGGAGCGCTTACGTGAGGAGGTATTCAGCTTATGACCTAAG 3905
QY 301 gGly-----IleLysGlnGlyIleTrpLysGluAlaMetSerAlaAspGlnAl 317
Db 3904 TCGTAGTCTTTCTTGACAGAGGCTGTGGAATGGAAGAC-----GAATATGC 3857
QY 317 aThrArgIleLysTyraAla-----SerLysTyraAlaGlnSerAlaAsnTyrrTrpLysAsnSe 336
Db 3856 CACTACCACTCAAGCAAGCAATGGGCAAGATGGGCTTCTGTATGACTATTCCTGTGAGCG 3797
QY 336 rIleGlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGlu 356
Db 3796 TTTCACTCTTATGATGAGTTTGTCA-----AAAGCTGTTCG 3761
QY 356 uArgAlaPheAlaAspTrpIleArgLys-----AsnGlyLysSerAlaValTy 372
Db 3760 TAAGTCTTTGTGAACCTTTCAGAAAGAGCTGATCATCGTGAGTATTCATCAAA 3701
QY 372 rGlyAsp-----ValLeuSerLeuGlnLysAlaTyrl 384
Db 3700 CTGGAGCCACACAGCTGCGACAGCCCTTCTGATATGAGGTGATTCACAAAGATAGAGA 3641
QY 385 -----GluGlyAlaLysAlaAsnArgLume 393
Db 3640 AGGTGCTTTCACACATGATTTACATGCTGGAAGATGTTACAGCGCTTGAAGTTGC 3561
QY 393 tThrTyrlLeuSerGluThrIleuPheGlyTyThrGluValAlaArgPheAlaGlnPheAl 413
Db 3580 TACAACCTGCTCGAGACTATGTTGGGAGCTT----- 3547
QY 413 aAsnAlaLeuAlaThrAsnProAspAlaHisAlaGlyIleLeuLysSerLeuAspArg 433
Db 3546 ---GCGGTTGCGGTTATTCAGAA-----GACCCGCG 3518

```


QY 433 sTyrLysasp-----TyrLeuProserLeuaspargLysValLeuPr 447
 :|||||
 Db 3517 CTACAGACCTTGATGGTAAAGCGTCATCTTCCA---ATCGTAATTAAGTCATCCC 3461
 QY 447 oAlaMetLeuasp-----IleValar 454
 :|||
 Db 3460 AATGCTTGAGATGAGCAGCAGATCTGAGCTTGACTGTGTCGTGCAAAATCACACC 3401
 QY 454 gArGArGTLleProAlaaspLysLeuProaspIlePheLysAsnValIleaspLysPhe 474
 :|||
 Db 3400 TGCCACAGCATCAATGACTCTTGTTGGCCAACTGATCACTGACCAAGTCACAGT 3341
 QY 474 eLysGLysPheThrLysLysTyrAlaaspPheValPheaspLysSerValProTyrSe 494
 :|||
 Db 3340 CATGACGACAGCAGCACTGATGAATGAGCTTGCTTGAATTTTCAGGCATG----- 3289
 QY 494 rAspLysPheHisAlaMetLeuLysSerMetaspLysGLysPheAlaLysAlaIleG1 514
 :|||||
 Db 3288 -GATCGTTTGAAGCTCGT-----AAGCAGTCGTGCTAAGTTGGAGA 3245
 QY 514 uLysaspProAlaValGluLeuSerLysSerValIleAlaAlaAlaArgAla----- 531
 :|||
 Db 3244 AATCGGTGCCCTTGCAAAATGCAAAACGTGTCACAGTGTGTCACAGAACGTAC 3185
 QY 532 -----IleG1 533
 :|||
 Db 3184 AGTGTCGTAATTGAGCCAGCTTGTCCTACTCAATGGTTCGTCAGAACGACCAATTGGC 3125
 QY 533 nAlaaspAlaMetAlaAsnAlaTyrAlaIleGluLysGLysArgLeuPhePheAlaG1 553
 :|||||
 Db 3124 TAAGAACCCATGCTGTAACCAAGACACAGAGACAG----- 3088
 QY 553 yLeuArgLysMetLysProGArgAlaLeuProser----- 565
 :|||
 Db 3087 ---GTCAATTGACCCAGCTGTTCAACGATACCTTCCTCAATGATGAAATGT 3032
 QY 566 ---AspAlaaspPheThrMetArgMetSerTyrGlySerIleLysGLyTyrGluProG1 584
 :|||
 Db 3031 CCAGCAGCGGGTTATCTCTGCTGACCTTGGTGGGTGAC-----CA 2990
 QY 584 nAspGlyAlaTyrPyrAsnTyrHisThrGlyLysGLyAlaLeuGluLysGlnAspPr 604
 :|||||
 Db 2989 AATCCCTCTGTCATCAAT-----GCTGATGCGTAATGATGATGCGGCAAGAACGTCC 2936
 QY 604 oLysSeraspGluPheAlaValGlnGluAsnIleLeuasp 617
 :|||
 Db 2935 AGAAGTGACGAGTGACTCAGAGACGAGACGCTTGAT 2896

RESULT 15
 AE007366 11280 bp DNA linear BCT 31-AUG-2001
 LOCUS Streptococcus pneumoniae TIGR4 section 49 of 194 of the complete
 DEFINITION genome.
 ACCESSION AE007366 AE005672
 VERSION AE007366.1 GI:14972031
 KEYWORDS
 SOURCE Streptococcus pneumoniae TIGR4.
 ORGANISM Streptococcus pneumoniae TIGR4.
 Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 Streptococcus.

REFERENCE
 1 (bases 1 to 11280)
 Tettelin, H., Nelson, K.E., Paulsen, I.T., Eisen, J.A., Read, T.D.,
 Peterson, S., Heidelberg, J., Deboy, R.T., Haft, D.H., Dodson, R.J.,
 Durkin, A.S., Gwinn, M., Kolonay, J.F., Nelson, W.C., Peterson, J.D.,
 Unayam, L.A., White, O., Salzberg, S.L., Lewis, M.R., Radune, D.,
 Holtzapple, E., Khouri, H., Wolf, A.M., Uterback, T.R., Hansen, C.L.,
 McDonald, L.A., Feldblyum, T.V., Angiuoli, S., Dickinson, T.,
 Hickey, E.K., Holt, I.E., Loftus, B.J., Yang, F., Smith, H.O.,
 Venter, J.C., Dougherty, B.A., Morrison, D.A., Hollingshead, S.K. and
 Fraser, C.M.
 Complete genome sequence of a virulent isolate of Streptococcus
 pneumoniae

JOURNAL Science. 293 (5529), 498-506 (2001)
 MEDLINE 21357209
 PUBMED 11463916
 2 (bases 1 to 11280)
 Tettelin, H., Nelson, K.E., Paulsen, I.T., Eisen, J.A., Read, T.D.,
 Peterson, S., Heidelberg, J., Deboy, R.T., Haft, D.H., Dodson, R.J.,
 Durkin, A.S., Gwinn, M., Kolonay, J.F., Nelson, W.C., Peterson, J.D.,
 Unayam, L.A., White, O., Salzberg, S.L., Lewis, M.R., Radune, D.,
 Holtzapple, E., Khouri, H., Wolf, A.M., Uterback, T.R., Hansen, C.L., McDonald, L.A.,
 Feldblyum, T.V., Angiuoli, S., Gesuwan, P., Hickey, E.K., Holt, I.E.,
 Loftus, B.J., Uwal, M.L., Yang, F., Smith, H.O., Venter, J.C.,
 Dougherty, B.A., Morrison, D.A., Hollingshead, S.K. and Fraser, C.M.
 Direct Submission
 TITLE Submitted (29-JUN-2001) The Institute for Genomic Research, 9712
 JOURNML Medical Center Dr, Rockville, MD 20850, USA
 FEATURES
 source
 1..11280
 /organism="Streptococcus pneumoniae TIGR4"
 /strain="TIGR4"
 /db_xref="taxon:170187"
 gene
 156..506
 /gene="SP0557"
 156..506
 /gene="SP0557"
 /note="identified by match to PFM protein family HMM
 PF02033"
 /codon_start=1
 /transl_table=1
 /product="ribosome-binding factor A"
 /protein_id="AAK74713.1"
 /db_xref="GI:14972032"
 /translation="MANHFRDVRGMEIKREVENIIOKKVRDPVVGTTIDYOMLGD
 LSAKYVYITLSNLSAQNKAQIGLEKATGTIRELGRNLKIKIPITLVKKESTIEY
 GKIDELRNLDKN"
 749..1564
 /gene="SP0558"
 749..1564
 /gene="SP0558"
 /note="identified by Glimmer2; putative"
 /codon_start=1
 /transl_table=1
 /product="conserved hypothetical protein"
 /protein_id="AAK74714.1"
 /db_xref="GI:14972033"
 /translation="MIRCKREISIVABODIAMOVLCQVVKLPYVVDKSKPLNDKF
 SKELDKDITPLEEQGPTTLISQETIDRVANACRDNVLAAGTSYLAGIPGLAMAI
 TTPADVAQFYAFSLKLAQELGYLYGEDIMASRESEDAQNLLLYLGVLMGVNCTA
 ALRVGSITITAKOYKIVPNKALTKLTPYLKVLKIFGVNLTGGLAGMGFTPI
 LGGIISGLTFEATKPMGESLQKELSKLVNVSEVOYQEDVETIRKEAIIKGE"
 1566..1748
 /gene="SP0559"
 1566..1748
 /gene="SP0559"
 /note="identified by Glimmer2; putative"
 /codon_start=1
 /transl_table=1
 /product="hypothetical protein"
 /protein_id="AAK74715.1"
 /db_xref="GI:14972034"
 /translation="MNPICAFAKIYGNPYLVGVKVMKTIKRAHVVGIGKLFAD
 KLMDTARMLIKPEERE"
 complement(1701..1934)
 /gene="SP0560"
 complement(1701..1934)
 /gene="SP0560"
 /note="identified by Glimmer2; putative"
 /codon_start=1
 /transl_table=1
 /product="hypothetical protein"
 /protein_id="AAK74716.1"
 /db_xref="GI:14972035"
 /translation="MEFLVLCNDYHLDFKPEIOYLKHYLVKQOLDCKRDDHPKEF
 HNPNNRDKRNSKTKKISFSLMLNPPSRIH"

```
gene
2196..2441
/gene="SP0561"
CDS
2196..2441
/feature="SP0561"
/feature="similar to GB:AE000782; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved domain protein"
/protein_id="AAK74717.1"
/db_xref="GI:14972036"
/translation="MEVMDNITDVSIPAEVVDKPEVLELIVELGFRPLANPLMRN
TVGRKVSILKQSGSLACTPMDKIVRTLEANGYEYIGLD"
2441..3775
/gene="SP0562"
CDS
2441..3775
/feature="SP0562"
/feature="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK74718.1"
/db_xref="GI:14972037"
/translation="MTDERIHILRDILLELHNGASPEVQDRPDATFTGSAIEISLM
BHEIAMSDSGVTFEDVWELCDVHANLFKNAIKGEVSDTEHPHPRVKEENLALRA
ALIRIRLADTYESMEDDEMLERKGLVRQMGIVQPDHQRKELEFPIMERGH
DSPPKVMGVVDQIRELQFALTATKSLPEVSISSYKAEFAATFESEMIPEESIL
LMILESFQDDMLQIAESDAVGAIIRPEKVPERSOFIEEKIAEPVOLDTAEV
QYOVQIDPEEGHTITFTPEKREAVIDRRSOAFGNGSVLQGANLILHLPDEITFV
NKEDITQYINDNTPADEMIFKRTPSQVGRNVELCHPKPLDKVTKIMKILRESKQKY
EMWFKESKSGKEVHTYAAVHDEGEGVLEVVDIQRREITDVFEGLE"
3785..4039
/gene="SP0563"
CDS
3785..4039
/feature="SP0563"
/feature="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK74719.1"
/db_xref="GI:14972038"
/translation="MSYEDPFMKFEFANVTOIMINDMAHKSQKYVEDDERAKDA
MIRIESRLDQILQKGFENFKVGKGFHDLPEGLFERNY"
4135..4530
/gene="SP0564"
CDS
4135..4530
/feature="SP0564"
/feature="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK74720.1"
/db_xref="GI:14972039"
/translation="MSKIKRKKQLNGLRRAGAPSSYTKVYVDENKRYKVAEASAS
AAGKAYSKVEQAVEATKEAQKANSVEDFANLGLDLDRAKFTYDGLIKASDFK
NMTEKELALKGIGPATIKRKENGIRKX"
4943..5392
/gene="SP0565"
CDS
4943..5392
/feature="SP0565"
/feature="similar to GP:8927331; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved domain protein"
/protein_id="AAK74721.1"
/db_xref="GI:14972040"
/translation="MIGVARENAEQIKOYKFTVINISDETSMAMEQAGFISHQEK
LGRIVGHEISERTQPIIDACPLVLDCEVDRIVEDGICHIFAKLIELRLVAPELLDE
KSHFKNQLFAPRYFMGDQRYRYRLDKRDKMSPIKARKKDKGN"
5379..5939
/gene="SP0566"
CDS
5379..5939
```

```
gene
/feature="SP0566"
CDS
/feature="acetyltransferase, GNAT family"
/codon_start=1
/transl_table=11
/product="acetyltransferase, GNAT family"
/protein_id="AAK74722.1"
/db_xref="GI:14972041"
/translation="MARTLPDKIETERLVLRVRYVDAEDIHITVASLPEVAPGFP
PVKILEDEITYLEHILPERTOKENLPAGYGIYKGDVIGSDPNHMLADVLELGY
TLHPDYMGNGYVPEAASLIDLAFFELGIHKITELCFYNLOSORSEKIGFTLEARI
RRKDOAGCCDDLRYGLRSEMEVY"
5924..6517
/gene="SP0567"
CDS
5924..6517
/feature="SP0567"
/feature="identified by match to TIGR protein family HMM
TIGR00422"
/codon_start=1
/transl_table=11
/product="valyl-tRNA synthetase"
/protein_id="AAK74723.1"
/db_xref="GI:14972042"
/translation="MGGLMHLFIIGAPASGKMTIGQELSRDLDTLFFYNQAIQFAL
EIVDYTEEMWEPFVGITPFSFLGASARNSRYLIDVIDFSNOYOLMYLKOIQLDN
PHOELFVELETSLERCIRNRTERNLKPKLRHIEVEREILFTAEITLQINSQHP
NELHHYFKINNTNLSMEAKQIQKINKIEGHTHV"
6510..9161
/gene="SP0568"
CDS
6510..9161
/feature="SP0568"
/feature="identified by match to TIGR protein family HMM
TIGR00422"
/codon_start=1
/transl_table=11
/product="valyl-tRNA synthetase"

Alignment Scores:
Pred. No.: 25.5 Length: 11280
Score: 129.00 Matches: 116
Percent Similarity: 33.62% Conservative: 77
Best Local Similarity: 20.21% Mismatches: 194
Query Match: 3.47% Indels: 187
DB: 1 Gaps: 26

US-10-008-355-2 (1-712) x AE007366 (1-11280)
QY 131 LysILEValLysValThrAspLysValGluGlyGlnLeuLysGlyIleThrAspGluMet 150
|||
6425 AAATATAATATAGCATTTGCTGCAGAAAGAGCTGCTACGAATTCATAATAATAA 6484
QY 151 GluArgLeu-ArgLysAlaGlnGluValCysGlnGluLeuAlaLysLysGlnLysAlaAs 170
|||||
6485 AACAAATATGACGAAGACACACACATGCTTAAAGAACTTACCTAAATACAAATCCAG 6544
QY 170 pGluAsnGlnLeuCysILEValGluProPheTyrSerAsnAsnGluTyrPheLeuIleVal 190
|||
6545 CGAGGTGGAGCGTGGTCTTACCAAAATGG-----CTTGA 6580
QY 190 IYrAspValPhe-----LysAspValArgMetValPheAl 202
|||||
6581 TGGGATGTTTCAAGCCTTCAGCGCATCAAAAGCTAAGCCTTATTCATCGTATTC 6640
QY 202 aProPheSerValGlyLysPheGlyGlyAspThrAspAsnTrpMetTrpProArgH 222
|||||
6641 ACCACCAACGTAACGTGGGAACCTT-----CA 6667
QY 222 sThrGlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgProAlaGluTyr 242
|||
6668 CCTTGGT-----CA 6676
QY 242 rSerLysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGlnGlyTyr 262
|||
6677 CGCTTGGATAGC---ACTTGCAGATATCATCATCCGTCACAAAACGATGCAAGCGTT 6733
QY 262 rLysAlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrLeu-- 281
|||
```

```
Db 6734 TGAT-----ACCTTGGCTTCAGGATGACGACCGCTGATTGC 6775
QY 282 -ThrserrgLyValGluAspArgIleGluAsnGluAsnProArgIleGluValAr 301
Db 6776 GACTCAGGCTAAGGTTGAGCGCTTACGCTGTGAGGGATTACGGTTATACCTAGG 6835
QY 301 gGly-----IleGlySlnGlyIleTrpLysGluAlaMetSerAlaSpGlnAl 317
Db 6836 TCGTGAAGTCTTCTTGACGAGGCTCGGGAATGGAAGAC-----GAATATGC 6883
QY 317 aThrArgIleLysTyrAla---SerLysTyrAlaGlnSerAlaSnTrpLysAsnSe 336
Db 6884 CACTACCATCAAGCAATAGGGGCAAGATGGGCTTCTGTAGACTATCTCTCGAGCG 6943
QY 336 rIleGlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGlu 356
Db 6944 TTTCACCTCTTGATGAAGGTTGTCA-----AAAGCTGTTCG 6979
QY 356 uArgAlaPheAlaAspTrpIleArgLys-----AsnGlyLysSerAlaValTyr 372
Db 6980 TAGGCTCTTGTGACCTTTACAGAAAGGCTGGATCTATCGTGTGATTCATCAA 7039
QY 372 rGlyAsp-----ValLeuSerSerLeuGluLysAlaTyrLys----- 384
Db 7040 CTGGGACCCAGCAGCTCGACAGCCCTTCTGATATTGAGTGATTCACAAGCATGAGA 7099
QY 385 -----GluGlyAlaLysAlaAsnArgLume 393
Db 7100 AGTGCCCTTACCACATGATTAACATCGTGAAGATGTTCAAGCCCTTGAACTTGC 7159
QY 393 tThrTyrLeuSerGluThrLeuPheGlyGlyThrGluValAlaArgPheAlaGlnPheAl 413
Db 7160 TACACACGCTCGTGAAGCTATGTTGGGAGCGT----- 7193
QY 413 aAsnAlaLeuAlaThrAsnProAspAlaHisAlaGlyIleLeuLysSerLeuAspLys 433
Db 7194 ----GCGGTTGCGGTTAATCCAGAA-----GACCCGCG 7222
QY 433 sTyrLysAsp-----TyrLeuProSerLeuAspArgLysValLeuPr 447
Db 7223 CTACAGAGGACTTGATGTTGTAATAAACGTCATCTTCCA--ATCGCTAATAAACATCATCC 7279
QY 447 oAlaMetLeuAsp-----IleValAr 454
Db 7280 AATCGTTGGAGATGAGACACCATCTGAGCTTGACTGCTGTAATAATCACACC 7339
QY 454 gArgArgIleProAlaAspLysLeuProAspIlePheLysAsnValIleAspLysLysPh 474
Db 7340 TGCCACAGATCCAATGACTTCTTGCGCAACGTCATACCTGCCACACAGTCACAGT 7399
QY 474 eLysGlyAspThrLysLysTyrAlaAspPheValPheAspLysSerValValProTyrSe 494
Db 7400 CATGAACGACGACGGAATGAAATGAGCTTGCTTGAATTTTCAGGCATG----- 7451
QY 494 rAspLysPheHisAlaMetLeuLysSerMetAspLysGluLysPheAlaLysAlaIleG 514
Db 7452 -GATCGTTTGAAGCTCGT-----AAGCAGCTGCTGCTAAGTTGGAGAGA 7495
QY 514 uLysAspProAlaValGluLeuSerLysSerValIleAlaAlaAlaArgAla----- 531
Db 7496 AATCGGTGCCCTTGTCAAAATCGAAAAACGTGCCACAGTGTGTCACACAGAACGTAC 7555
QY 532 -----IleG 533
Db 7556 AGGTGCTAGTGTGAGCCAGCTTGTCTACTCAATGGTTCGTCACAAGATGAGCAATTTGC 7615
QY 533 nAlaAspAlaMetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPhePheAlaG 553
Db 7616 TAAGAAAGCCATGCTTACCAAGACACAGAGACAG----- 7652
QY 553 yLeuArgGluMetLysProGlyArgAlaLeuProSer----- 565
Db 7653 ----GTGGAATTTACCCACCTGTTTCAACGATACCTTCTTCATGATGGAATAATGT 7708
```

```
QY 566 ----AspAlaAsnPheThrMetLArgMetSerTyrGlySerIleLysGlyTyrGluProG 584
Db 7709 CCAAGACTGGCTTATCTCTGTCAGCTTGTGTGGGTCAC-----CA 7750
QY 584 nAspGlyAlaTrpLysAsnTyrHisThrThrGlyLysGlyValLeuGluLysGlnAspPr 604
Db 7751 AATCCCTGCCCTGGTACAAAT-----GCTGATGGTGAATAATGTATGTGCGGACAGAGCTCC 7804
QY 604 oLysSerAspGluPheAlaValGlnGluAsnIleLeuAsp 617
Db 7805 AGAAGTGAAGATGAGACTCAGACGAAAGAGCTTGGAT 7844
```

Search completed: October 17, 2002, 23:36:03
Job time : 2700 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 17, 2002, 22:39:21 : Search time 217 Seconds

(without alignments)
5633.378 Million cell updates/sec

Title: US-10-008-355-2

Perfect score: 3719

Sequence: 1 MOKKLSILGALLGASG.....LFMIDKMGCCPRILIGELKLI 712

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1736436 segs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+g2n.model -DEV=xlh
-O=/cgn2_1/USPRO.spool/US10008355/runat_16102002_121152_27685/app_query.fasta_1.903
-DB=N.Geneseq_032802 -QFWT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.ccd1
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFWT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10008355 -ECGN_1_1_65_etunal_16102002_121152_27685 -NCPV=6 -ICPV=3
-NO_XLPHY -NO_MMAP -LARGEDUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCGAPOP=6 -FCGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N.Geneseq_032802: *
1: /SIDSL1/gcgdata/hold-geneseq/geneseqn-emb1/NA1980.DAT: *
2: /SIDSL1/gcgdata/hold-geneseq/geneseqn-emb1/NA1981.DAT: *
3: /SIDSL1/gcgdata/hold-geneseq/geneseqn-emb1/NA1982.DAT: *
4: /SIDSL1/gcgdata/hold-geneseq/geneseqn-emb1/NA1983.DAT: *
5: /SIDSL1/gcgdata/hold-geneseq/geneseqn-emb1/NA1984.DAT: *
6: /SIDSL1/gcgdata/hold-geneseq/geneseqn-emb1/NA1985.DAT: *
7: /SIDSL1/gcgdata/hold-geneseq/geneseqn-emb1/NA1986.DAT: *
8: /SIDSL1/gcgdata/hold-geneseq/geneseqn-emb1/NA1987.DAT: *
9: /SIDSL1/gcgdata/hold-geneseq/geneseqn-emb1/NA1988.DAT: *
10: /SIDSL1/gcgdata/hold-geneseq/geneseqn-emb1/NA1989.DAT: *
11: /SIDSL1/gcgdata/hold-geneseq/geneseqn-emb1/NA1990.DAT: *
12: /SIDSL1/gcgdata/hold-geneseq/geneseqn-emb1/NA1991.DAT: *
13: /SIDSL1/gcgdata/hold-geneseq/geneseqn-emb1/NA1992.DAT: *
14: /SIDSL1/gcgdata/hold-geneseq/geneseqn-emb1/NA1993.DAT: *
15: /SIDSL1/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT: *
16: /SIDSL1/gcgdata/hold-geneseq/geneseqn-emb1/NA1995.DAT: *
17: /SIDSL1/gcgdata/hold-geneseq/geneseqn-emb1/NA1996.DAT: *
18: /SIDSL1/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT: *
19: /SIDSL1/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT: *
20: /SIDSL1/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT: *
21: /SIDSL1/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT: *
22: /SIDSL1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT: *
23: /SIDSL1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT: *
24: /SIDSL1/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	134.5	3.6	2853	20 AAX91754	Porphyromonas ging
2	134.5	3.6	2859	20 AAX91621	Porphyromonas ging
3	133	3.6	2652	23 AAS55647	Streptococcus pneu
4	130	3.5	2652	20 AAX37027	S. pneumoniae vail
5	130	3.5	2652	20 AAX32814	S. pneumoniae vail
6	130	3.5	2652	21 AAA14369	DNA encoding Strep
7	129	3.5	6004	19 AA052160	Streptococcus pneu
8	127	3.4	2652	23 AAS55949	Streptococcus pneu
9	123.5	3.3	2142	20 AA231948	M. catarrhalis str
10	119.5	3.2	3271	22 AA164839	KRN-100 coding seq
11	117.5	3.2	1664976	19 AA420203	Methanococcus jann
12	117	3.1	1830121	17 AA420203	Haemophilus influe
13	116.5	3.1	15424	21 AA420203	Porcine lelystad v
14	116.5	3.1	15424	21 AAC81764	Porcine lelystad v
15	116	3.1	3598	21 AAC77776	Human cancer assoc
16	115	3.1	4853	17 AAT13139	Alpha-D-glucosyltr
17	114	3.1	2136	17 AAT11244	Neisseria meningit
18	114	3.1	3420	22 AAH14452	Human cDNA sequenc
19	114	3.1	3431	21 AAC76951	Human OREF ORF2506
20	113.5	3.1	2856	23 AAS52731	E. coli DNA for ce
21	113	3.0	8930	19 AA022834	Haemophilus paraga
22	112.5	3.0	36471	19 AA081453	N. meningitidis pa
23	112.5	3.0	349980	21 AAF21611	Neisseria meningit
24	112.5	3.0	837096	21 AA081489	N. meningitidis pa
25	112	3.0	2175	23 AAS79571	DNA encoding novel
26	112	3.0	2175	23 AAS81718	DNA encoding novel
27	112	3.0	3285	22 AAH54398	S. epidermidis gen
28	112	3.0	56506	21 AA069168	Bacteriophage Dp-1
29	111	3.0	4411529	22 AA199682	Mycobacterium tube
30	110.5	3.0	3279	14 AA051556	Sequence encoding
31	110.5	3.0	12665	19 AAV52267	Streptococcus pneu
32	110.5	3.0	15450	21 AAA27809	C glutamic acid por
33	110.5	3.0	349980	22 AAH68530	Nucleotide sequenc
34	110.5	3.0	1230025	20 AAX91990	Neisseria meningit
35	110	3.0	2255	17 AAT11240	Thermotoga maritim
36	110	3.0	2319	19 AAV36920	C. symbiosum open
37	110	3.0	2538	21 AA055217	Sequence encoding
38	110	3.0	3279	14 AA050946	Cenarchaeum symbio
39	110	3.0	32998	21 AA055186	Streptococcus pneu
40	109.5	2.9	2082	21 AA050408	Thermotoga maritim
41	109.5	2.9	2319	18 AAT93691	Human secreted pro
42	109	2.9	1490	21 AA059611	Streptococcus pneu
43	109	2.9	6703	21 AA050436	Streptococcus pneu
44	109	2.9	15363	19 AAV52272	Pyrococcus abyssi
45	109	2.9	349980	22 AAH41225	

ALIGNMENTS

RESULT 1
AAX91754
ID AAX91754 standard; DNA; 2853 BP.

XX AAX91754;

DT 25-AUG-1999 (first entry)

XX Porphyromonas gingivalis protein Pg67 encoding DNA.

DE Porphyromonas gingivalis; PG; periodontal disease; gingivitis;

KW vaccine; antigenic; ds.

OS Porphyromonas gingivalis.

XX WO9929870-A1.

PN 17-JUN-1999.

XX 10-DEC-1998; 98MO-AU01023.
 XX 04-AUG-1998; 98AU-0005028.
 PR 10-DEC-1997; 97AU-0000839.
 PR 31-DEC-1997; 97AU-0001182.
 PR 30-JAN-1998; 98AU-0001546.
 PR 10-MAR-1998; 98AU-0002264.
 PR 09-APR-1998; 98AU-0002911.
 PR 23-APR-1998; 98AU-0003128.
 PR 05-MAY-1998; 98AU-0003338.
 PR 22-MAY-1998; 98AU-0003654.
 PR 29-JUL-1998; 98AU-0004917.
 XX (CSLC-) CSL LTD.
 XX Agius CT, Barr IG, Hocking DM, Margetts MB, Patterson MA;
 PI Ross BC, Rothel LT, Webb EA;
 XX WPI, 1999-385613/32.
 DR P-PSDB; AAY34536.
 XX Antigenic Porphyromonas gingivalis peptides for preventing
 PT gingivitis
 PS Claim 12; Page 244-245; 588pp; English.
 XX AAX91536 to AAX91801 encode two hundred and sixty six antigenic
 CC Porphyromonas gingivalis (PG) polypeptide sequences given in AAY34318 to
 CC AAY34583. AAX91802 to AAX91989 represent PCR primers used in the
 CC isolation of the PG polypeptides. The PG polypeptides have antibacterial
 CC activity with a vaccine mechanism of action. The PG polypeptides can be
 CC used as vaccines especially against Porphyromonas gingivalis. Probes can
 CC be used to detect Porphyromonas gingivalis in standard hybridisation
 CC assays. Porphyromonas gingivalis is involved in periodontal disease
 CC especially gingivitis.
 XX
 SQ Sequence 2853 BP; 736 A; 685 C; 691 G; 741 T; 0 other;
 Alignment Scores:
 Pred. No.: 0.00723 Length: 2853
 Score: 134.50 Matches: 105
 Percent Similarity: 34.41% Conservative: 66
 Best Local Similarity: 21.13% Mismatches: 202
 Query Match: 3.62% Indels: 124
 DB: 20 Gaps: 23
 US-10-008-355-2 (1-712) x AAX91754 (1-2853)
 QY 221 ArgHisThrGlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArg-ProAl 240
 DB 717 CGGTCACAGTCGCGTCACATCTGCTTATCGCCGATATGCCGTCGCGATAGTCTTCTT 776
 QY 240 aglUtySerLysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetG1 260
 DB 777 CGGCTATTTT-----CCCTTTCCACAAATACTCTTCGCTATTTGATGCC 824
 QY 260 nglTYTYLysAlaAspAspTyrAlaMetThrLleGlyPheProGlySerThrAspArgTY 280
 DB 825 CAGGTACGAGAGGACAT-----CGCTATGATTT-----TA 857
 QY 280 rLeuThrSerTyrPgly-----ValGluAspArgLleGluAsnGluAsnAsnPr 296
 DB 858 TTTCAGGAAATGCGATATATTTTTCCTTCAGGACTATATCGAT----- 903
 QY 296 oArgLleGluValArgLly-----LleYsGlnGlyLleTyrLysGluAlaMetSerAl 314
 DB 904 -----TTGGCATTCGCGGAGATCTTTTCCAAAGGTCATTCACGCCCATC 959
 QY 314 aAspGlnAlaThrArgLleLysTyrAlaSerLysTyrAlaGlnSerAlaSnTyrTply 334
 DB 960 GAAATATACAGAGGTATATATGATACACGCGCTGCTC-----GAAACCAATTA 1008

QY 334 sAsnSerLleGlyMetAsnArgLysLeuAlaArgLeuAspValLleGlyArgLysArgAl 354
 DB 1008 ----- 1008
 QY 354 aglUgluArgAlaPheAlaAspTrrPleArgLysAsnGlyLysSerAlaValTyrGlyAs 374
 DB 1009 -----CTGTATCGAAAGTCCGGGACAAATATCTGCGCGAGA 1046
 QY 374 rValLeuSer-----SerLeuGlnLysAlaTyrLysGlnLysAlaAs 390
 DB 1047 CTACAGCAACACCCAGCTGCTGAATATCCATGAGACACACAGTCAGATCCGAAGCCAA 1106
 QY 390 nArgGluMetThrTyrLeuSerGlyLysThrLeuPheGlyGlyThrGluValArgPheAl 410
 DB 1107 TCCT-----TTGCAACGTTGTGCGCAATGTCAAATTTTCCACCGGAG 1151
 QY 410 aglInPheAlaAspAlaLeuAlaThrAsnProArgAlaHisAla-----GlyLleLeuY 428
 DB 1152 CTATTTCCAGAAATTCGCTGAATTCACCACTATGATGTCATATGCGTACTGTCAGACAG 1211
 QY 428 sSerLeuAspAspLysTyrLysAspTyrLeuProSerLeuAspArgLysValLeuProAl 448
 DB 1212 AAGTTCGCGCGGTAGAGCTATTCGCGCAAGTTCCGGGTACTCTTTTCGATTCAGGCTAG 1271
 QY 448 aMetLeuAspLle-----ValArgArgArgLleProAlaAspLysLeuProAspAl 465
 DB 1272 CATG---GATATAGCCGAGACATGCGCGATACAGCGGTAGCTTACTCTTGGCGAATCT 1328
 QY 465 ePheLysAsnValLleAspLys---LysPheLysGlyAspThrLysTyrLysAlaSprh 484
 DB 1329 TTGCATTAATATATGTCACGCGGTATTCCTTTCACAGGACACCGGCTAGACCGGCGC 1388
 QY 484 eValPheAspLysSerValAlProTyrSerAspLysPheHisAlaMetLeuLysSerMe 504
 DB 1389 ATGTAGACAGAAAGTGTAGTGTGGGTATTCGCGACAGCTCGCAATAGTATCTTGACAA 1448
 QY 504 tAspLysGlnLysPheAlaLysAlaLleGluLysAspProAlaValGluLeuSerLysSe 524
 DB 1449 AGAAGAAAGATTCTCCAGACATCTCGTCCGATGGAGAAATGCTTGGCTATTC 1508
 QY 524 rValLleAlaAlaAlaArgAlaLleGlnAlaAspAlaMetAlaAsnAlaTyrAlaLleG1 544
 DB 1509 CGTACCGATAGTTGACTCTCCCTTGTGATATATATACATGACTATGGGGGTAA 1568
 QY 544 uLysGlyLysArgLeuPhePheAlaGlyLeuArgGluMetTyr-----ProGlyAr 561
 DB 1569 CTACAAATGAGTGGGTGATACAGAAAGCATATCGGATGGAGATGATGAGATAAGAAAC 1628
 QY 561 gAlaLeuProSerAspAlaAsnPheThrMetArgMetSerTyrGlySerLleLysGlyTy 581
 DB 1629 ATTCTGCTCTTCGAGACGACGACATATAATTCGCGACAGCTGAC-----GATTA 1676
 QY 581 rGluProGlnAspGlyAlaTrrTyrAsnTyrHisThrGlyLysGlyValLeuGluLyu 601
 DB 1677 CAGTCTGTCCGACAGC-----TTATGTACACATTTGTACGGTATGTTCAAGCC 1724
 QY 601 sGlnAspProLysSerAspLysPheAlaValGlnGluAsnLleLeuAspLeu----- 618
 DB 1725 TTGGAACCT-----TTTTCCTCGGAGGAAATCTGATATGATCGGTATCG 1772
 QY 619 -----PheArgThrLysAsnTyrGlyAr 636
 DB 1773 CTTCACGCCCATCTGAGTTTCTCTATATGCGCGATCTACAAACGCGCATGAGCTT 1832
 QY 626 gTyrAla-----GluAsnGlyGlnLleHisLleAlaPheLeuSerAs 640
 DB 1833 TTGGAGACTTCTTGAGCATACGAGATCAGACAGCGACAGTCGATACGCTCTACTCTCC 1892
 QY 640 n-----AsnAspLleThrGly-----GlyAsnSerGlySerProValPheAs 654
 DB 1893 TTATTTTCAGACAAATATTCGCTGCTCCCTCATAGGCAATGCAAGATCT----- 1941
 QY 654 pLysAsnGlyArgLeuLleGlyLeuAlaPheAspGlyAsnTrrPgluAla 670

Db 1942 -----:::|||||
RESULT 2
AAK91621
ID AAK91621 standard; DNA; 2859 BP.
AC AAK91621;
XX
XX
XX 25-AUG-1999 (first entry)
DE Porphyromonas gingivalis protein PG67 ORF encoding DNA.
XX
XX Porphyromonas gingivalis; PG; periodontal disease; gingivitis;
KM vaccine; antigenic; ds.
XX
XX Porphyromonas gingivalis.
PN WO9929870-A1.
XX
XX 17-JUN-1999.
PD
XX
XX 10-DEC-1998; 98WO-AU01023.
PF
XX
XX 04-AUG-1998; 98AU-0005028.
PR 10-DEC-1997; 97AU-0000839.
PR 31-DEC-1997; 97AU-0001182.
PR 30-JAN-1998; 98AU-0001546.
PR 10-MAR-1998; 98AU-0002264.
PR 09-APR-1998; 98AU-0002911.
PR 23-APR-1998; 98AU-0003128.
PR 05-MAY-1998; 98AU-0003338.
PR 22-MAY-1998; 98AU-0003654.
PR 29-JUL-1998; 98AU-0004917.
XX
XX (CSLC-) CSL LTD.
XX
XX Agius CT, Barr IG, Hocking DM, Margetts MB, Patterson MA;
PI Ross BC, Rothel LJ, Webb EA;
XX
XX WPI; 1999-385613/32.
DR P-PSDB; AAY34403.
XX
XX Antigenic Porphyromonas gingivalis peptides for preventing
PT gingivitis
XX
XX Claim 12; Page 153-154; 588pp; English.
PS
XX
XX AAK91536 to AAK91801 encode two hundred and sixty six antigenic
CC Porphyromonas gingivalis (PG) polypeptide sequences given in AAY34318 to
CC AAY34583. AAK91802 to AAK91989 represent PCR primers used in the
CC isolation of the PG polypeptides. The PG polypeptides have antibacterial
CC activity with a vaccine mechanism of action. The PG polypeptides can be
CC used as vaccines especially against Porphyromonas gingivalis. Probes can
CC be used to detect Porphyromonas gingivalis in standard hybridisation
CC assays. Porphyromonas gingivalis is involved in periodontal disease
CC especially gingivitis.
XX
XX
SQ Sequence 2859 BP; 738 A; 687 C; 692 G; 742 T; 0 other;

Alignment Scores:
Pred. No.: 0.00725 Length: 2859
Score: 134.50 Matches: 105
Percent Similarity: 34.418 Conservative: 66
Best Local Similarity: 21.138 Mismatches: 202
Query Match: 3.628 Indels: 124
DB: 20 Gaps: 23

US-10-008-355-2 (1-712) x AAK91621 (1-2859)
QY 221 ArgHStHrGlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArg-ProAl 240
DB 723 CGTCAAGGTCGGTCATCTGTATTCGCCGATATGCGCGTGCATGAGTCTTCTT 782

QY 240 aGluTyrSerIysAspAsnLysProTyrIysProValTyrPheAlaAlaValSerMetC1 260
DB 783 CCGCTATTTT-----CCTTTCCACAATATCTTCCGATATTTGATGCC 830
QY 260 nGlyTyrLysAlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTy 280
DB 831 CACGTACGGAGAGACAAT-----CGCTATGGATT-----TA 863
QY 280 rLeuThrSerTyrGly-----ValGluAspArgIleGluAsnGluAsnAsnPr 296
DB 864 TTTGAGGAATGGTGATATATTTTTCCTTCAGCGCATATACAT----- 909
QY 296 oArgIleGluValArgGly-----IleLysGlnGlyIleTyrLysGluAlaMetSerAl 314
DB 910 ---TTGGCATTCGTTGGAGATCTTTTCCAAAGGTCATGGGCGCATTTCCGCCAATC 965
QY 314 aAspGlnAlaThrArgIleLysTyrAlaSerLysTyrAlaGlnSerAlaAsnTyrTyrLy 334
DB 966 GAAATATAGAAGAGGTATATAGTACACGCTCTTC-----GAAGCCAATTAT----- 1014
QY 334 sAsnSerIleGlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAl 354
DB 1014 ----- 1014
QY 354 aGluGluAlaArgAlaPheAlaAspTyrIleArgLysAsnGlyLysSerAlaValTyrGlyAs 374
DB 1015 -----CTGATACGAATCCGCGCAATAATACCGCGGAGA 1052
QY 374 pValLeuSer-----SerLeuGluLysAlaTyrLysGluGlyAlaLysAlas 390
DB 1053 CTACAGCAAGACACACAGCTGATATCCGATGAGCACACAGTACAGATCCGAAGCCAA 1112
QY 390 nArgGluMetThrTyrLeuSerGluThrLeuPheGlyIleThrGluValValArgPheAl 410
DB 1113 TCCT-----TTGCAACGTTGTGCGCCATGCAATTTTGCACCGGAG 1157
QY 410 aGlnPheAlaAsnAlaLeuAlaThrAsnProAspAlaHisAla-----GlyIleLeuLy 428
DB 1158 CTATTTCCAGAAATGCTGATATACCACTATGATGATCAATGACCGCTACTGACAGACAG 1217
QY 428 sSerLeuAspAspLysTyrLysAspTyrLeuProSerLeuAspArgLysValLeuProAl 448
DB 1218 AAGTTGCGCGGTGAGCTATTCGCGCAAGTTTCCGGTACTCTTTTTCATTAACGGGTAG 1277
QY 448 aMetLeuAspIle-----ValArgArgArgIleProAlaAspLysLeuProAspI1 465
DB 1278 CATG---GATATCAGCCAGACATGCGCATGACAGCGTACCTTACCTGCGCATCT 1334
QY 465 ePheLysAsnValIleAspLys---LysPheLysGlyAspThrLysLysTyrAlaAspPh 484
DB 1335 TTCGATTAATATGTCACCGCTTATCTTTCAAGCGGAAGCCCGTGTGACCGGAGCG 1394
QY 484 eValPheAspLysSerValValProTyrSerAspLysPheHisAlaMetLeuLysSerMe 504
DB 1395 ATGGRACGGAAGTTGAGTGTGGCTATTCGGTCACTTCCCATATAGTATCTTGACAA 1454
QY 504 tAspLysGluLysPheAlaLysAlaIleGluLysAspProAlaValGluLeuSerLysSe 524
DB 1455 AGAGAAAGATTGTGCCACAGACATCTCGTGGCATTTGAGAAATGATGATCGCATATC 1514
QY 524 rValIleAlaAlaAlaArgAlaIleGlnAlaAspAlaMetAlaAsnAlaTyrAlaIle1 544
DB 1515 CGTACCGATCACTTTGACTGCTCTTTGTGGATATATCAATCTGACTATGAGGGGTAA 1574
QY 544 uLysGlyLysArgLeuPheAlaGlyLeuArgGluMetTyr-----ProGlyAr 561
DB 1575 CTACATATAGTGTGTATACGAAGGCATACGAAGTGTGATGAGATTAAGAAAC 1634
QY 561 gAlaLeuProSerAspAlaAsnPheThrMetArgMetSerTyrGlySerIleLysGlyTy 581
DB 1635 ATTCTGCTTGGACACAGCACTATTAATTCGACAGACTGTAC-----GATTA 1682

Oy	581	rgluproglinsplalatrphyriasmnrhstlrthrlgylslgvallleuglyl	601
Dd	1663	cagctgtgcggcaggc-----ttatctaccacattgrracggatcttccaaGCC	1730
Oy	601	sglnasprolysSeraspeluphealaValngluuasnilleusapLeu-----	618
Dd	1731	TTGGAAACCT-----TTTTCCITCGGAGCAATTCATTATGCATCCTGCATCG	1778
Oy	619	-----PheatgThlyAsnTyrGLYAr	626
Dd	1779	CTTCACGCCCATCTCAGTTTCCTCATAAGCCGGAACCTCACGAAGAAGCCGATATGGCTT	1838
Oy	626	gTYrAla-----gluasngilylnileuhisilealpheleuSerAs	640
Dd	1839	TTGGGAGGCTTTTGAGCATACGGATCAGAAAGCGAAGCTGACGTCTACTCTCC	1898
Oy	640	n-----AsnaSplerhrclyl-----glyasnserglySerProvalPheas	654
Dd	1899	TTATTTGAGAGCATATTTCGTCGTCCTCCATCGGACGATGCAGATCT-----	1947
Oy	654	plyasngilyarLeuileclyleualaPheaspglyasnTprgluala	670
Dd	1948	-----GTCAAATTCTCTTTGGACAACCAACTAGAGGCC	1980

RESULT 3
 AAS55647
 ID AAS55647 standard; DNA; 2652 BP.
 XX
 AC AAS55647;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE Streptococcus pneumoniae DNA for cellular proliferation protein #218.
 XX
 KW Antisense; ds; prokaryotic cellular proliferation gene;
 XX antibiotic; antibacterial; drug design.
 XX
 OS Streptococcus pneumoniae.
 XX
 PN WO200170955-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 21-MAR-2001; 2001WO-US09180.
 XX
 PR 21-MAR-2000; 2000US-191078P.
 XX 23-MAY-2000; 2000US-206848P.
 PR 26-MAY-2000; 2000US-207727P.
 XX 23-OCT-2000; 2000US-242578P.
 PR 27-NOV-2000; 2000US-253625P.
 XX 22-DEC-2000; 2000US-257931P.
 PR 16-FEB-2001; 2001US-269308P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr CJ;
 PI Yamamoto RT, Xu HH;
 XX
 XX WPI: 2001-611495/70.
 DR P-PSDB; AAU37788.
 XX
 PT New polynucleotides for the identification and development of
 XX antibiotics, comprise sequences of antisense nucleic acids -
 PS
 Claim 27; Seq ID No 9284; 511pp; English.
 XX
 XX The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the
 CC genes, their use in the discovery of novel antibiotics, the essential
 CC genes themselves and the encoded proteins. The prokaryotes used are
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
 CC pneumoniae, pseudomonas aeruginosa and Enterococcus faecalis. The
 CC invention is also useful for the identification of potential new targets

for antitoxic development. The antisense nucleic acids can also be used to identify proteins used in proliferation, to express these proteins, and to obtain antibodies capable of binding to the expressed proteins. The proteins can be used to screen compounds in rational drug discovery programmes. The antisense nucleic acid sequence is also useful to screen for homologous nucleic acids which are required for cell proliferation in a wide variety of organisms. The present sequence encodes an essential prokaryotic cellular proliferation protein.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences).

Alignment Scores:

Pred. No.:	0.00913	Length:	2653
Score:	133.00	Matches:	106
Percent Similarity:	33.20%	Conservative:	64
Best Local Similarity:	20.70%	Mismatches:	164
Query Match:	3.58%	Indels:	178
DB:	23	Gaps:	25

US-10-008-355-2 (1-712) x AAS55647 (1-2652)

QY	192	AspValPhe-----	-----LysAspValArgMetValPheAlaPro	203
Db	76	GATGTTTTCAGGCCCTTCAGCGGATCAAAAGCCTTAAGCTTAATTCACGTTATTCACCA	135	
QY	204	ProSerSerValGlyLysPheGlyGlyLysPheTrpAsnTrpMetTrpProArgHisThr	223	
Db	136	CCAAACGTTACAGGAAACTT-----	-----CACCTT	162
QY	224	GlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgProAlaGluTyrSer	243	
Db	163	GGT-----	-----CACGCT	171
QY	244	LysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGlnGlyTyrLys	263	
Db	172	TGGGATACA---ACTTTCAGAGATATTATTCATCCGTCGCAAAACCCATGACAGAGTTTGAT	228	
QY	264	AlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrLeu---Thr	282	
Db	229	-----ACCCTTGGCTCTCTGGATGACACACGAGGAGATTGCCACT	270	
QY	283	SerTrpGlyValGluAspArgIleGluAsnGluAsnAsnProArgIleGluValArgGly	302	
Db	271	CAGGCTAAGGTTGAGAGACCGCTTGCGTGTAAGGCGCATTTCCCGCTATGACCTTGGCTG	330	
QY	303	-----IleLysGlnGlyIleTrpLysGluAlaMetSerAlaAspGlnAlaThr	318	
Db	331	GAGCTTTCCTGACCAAAAGTCGGGAATGGAAAGC-----	-----GAATAGCCACT	378
QY	319	ArgIleLysTyrAla---SerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsnSerIle	337	
Db	379	ACCATCAAGGAACAATGGGGCAGAGTGGCGCTCTGTAGACATTTCTCGTGAACCGTTTC	438	
QY	338	GlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGluGluArg	357	
Db	439	ACTCTTGACGAAGGTTTCTCA-----	-----AAAGCTGTTCCGTAAG	474
QY	358	AlaPheAlaAspTrpIleArgLys-----	-----AsnGlyLysSerAlaValTyrGly	373
Db	475	GTCCTTGTGGACCTTACAAAGAAAGCGATGATACCGGTGGAGTTATCATCAACAACGTG	534	
QY	374	Asp-----ValLeuSerSerLeuGlnLysAlaTyrLys-----	-----	384
Db	535	GACCACGAGCGTCGCACACGCCCTTCTGATGTAGAGTGATTCACAAAGATGTGGAAGGT	594	
QY	385	-----GlnGlyAlaLysAlaAsnArgGluMetThr	394	
Db	595	GCCCTTACACATGAATTACATGCTGGAAGATGATGGTTCACAGCGCCCTTGAGTTGCTTACA	654	

Oy	374	ASP-----ValLeuSerLeuGluLysAlaIleTyrGlySGLuGlyLysAlaLys	388
Dd	535	GACCACGAGCTGGCACACGCCCTTTCTGTATATTGAAGTGATTCACAGAAGTGTGAAGT	594
Oy	389	AlAsnArgrGlumetThrTyLeu-----	396
Dd	595	GCGCTTACCAACATGATTTACATGCTGGAAAGATGGTTACAGCGTCTGTAACTGGTACA	654
Oy	397	-----SeriLuthrLeupheGlyGlyThrGluValValArgPheAlaGlnPheAlaasn	414
Dd	655	ACTGCTCCTGTAGAACCATGTTGGGGAGCTT-----	684
Oy	415	AlAleuAlaThrAsnProAspAlaHisAlaGlyLileuLysSerLeuaspasplyTyr	434
Dd	685	GCGGTGGCGTCAACCCAGAA-----GACCCGGGCTAC	717
Oy	435	Lysasp-----TyrLeuproSerLeuAsparGlySvalleuProAla	448
Dd	718	AAGGACTTGATGTATAAATATGCATCCTTCCA---ATCGCTAATTAACATCATCCCAATC	774
Oy	449	Meteuasp-----lleValArgarg	455
Dd	775	GTTGGAGATGAGCAGCAGATCCTGAATTTGTACTGGTGGTGGTGAANAATCACACCTGCC	834
Oy	456	ArgLleProAlaAsplylsleuProAspLlePheLysAsnValLleasplylsypheLys	475
Dd	835	CACGATCCAAATGACTTTCTGGTGGCCCAACGTATCACTGGCCACACAGTCAACGTATG	894
Oy	476	GlyAspThrLysLysTyrAlaAspPheValPheAspLysSerValValProTyrSerasp	495
Dd	895	AACGACGACGGAACCATGATGACTTGCGCTTGTGATTTCAACGCATG-----GAC	945
Oy	496	LysPheHisAlaMetLeuLysSerMetAspLysGluLysPheAlaLysAlalleuLys	515
Dd	946	CGTTTTGAAAGCTCT-----AAGCAGCTGGTGGCTAACTGGAAGAAAATC	990
Oy	516	AspProAlaValGluLeuSerLysSerValLleAlaAlaAlaArgla-----	531
Dd	991	GGTGCCCTCGTCAAATCGAAAAACGTGTCCACAGTGTGGTGCATCTACAGCGTACAGST	1056
Oy	532	-----lleGlnAla	534
Dd	1051	GTTGTGGTTAAACCTCGCTGTCTACTCAATGGTTGCGTCAAGATGACCAATTTGGCTAAG	1116
Oy	535	AspAlaMetAlaAsnAlaTyrAlaAlleGluLysGlyLysArgLeuPheAlaGlyLeu	554
Dd	1111	AACGCGATTTGGCAACAGACACAGAGAGACAG-----	1143
Oy	555	ArgGluMetTyrProGlyArgAlaLeuProSer-----	565
Dd	1144	GTCGAATTTTACCCACCTCGTTTCAACGATACCTTCTTCATATGATGATGAAAAATGCCAC	1201
Oy	566	AspAlaAsnPheThrMetArgMetSerTyrGlySerIleLysGlyLysIleuProGlnasp	585
Dd	1204	GACGGGGTTATCTCTGTCAGCTCTGTGGTGGGCTCAC-----CAAAATC	1246
Oy	586	GlyAlaTrrpyrAsnTyrHisThrThylGlyLysGlyValleuLysLysInsProLys	605
Dd	1246	CTCTCCCTGTACAT-----GCTGATGATGGAATGTATGTGGCGAAMAACTCCACAGA	1299
Oy	606	SerAspGluPheAlaValGlnGluAsnIleLeuasp	617
Dd	1300	GGTGACGAGATGACTCAGACAGCAAGACGTTTGGAT	1335
RESULT	5		
AXX	AAK32814		
ID	AAK32814	standard; DNA; 2652 BP.	
XX	AAK32814;		
DT	25-JUN-1999	(first entry)	

DE	S. pneumoniae valyl tRNA synthetase (vals) polypeptide encoding DNA.
XX	
KM	Valyl tRNA synthetase; vals; otitis media; conjunctivitis; pneumonia;
KM	bacteremia; sinusitis; pleural empyema; endocarditis; meningitis;
KM	bacterial adhesion; bacterial infection; ss.
XX	
OS	Streptococcus pneumoniae.
XX	
PN	EP009818-A1.
XX	
PD	21-Apr-1999.
XX	
PF	17-OCT-1997; 97EP-0308257.
XX	
PR	17-OCT-1997; 97EP-0308257.
XX	
PA	(SMIK) SMITHKLINE BEECHAM CORP.
XX	
PI	Brown JR, Jaworski DD, Lawlor EJ, Wang M;
DR	WPI: 1999-231674/20.
XX	P-PSDB: AAY06828.
XX	
PT	New Streptococcus pneumoniae valyl tRNA synthetase (vals) gene and
PT	protein, useful for screening for antibacterial compounds, for
PI	prevention and treatment of Streptococci infections e.g. meningitis
PT	and conjunctivitis
XX	
PS	Claim 4; Page 6-9; 35pp; English.
XX	
CC	This DNA encodes a S. pneumoniae vals polypeptide (EC 6.1.1.9), a mem-
CC	ber of the valyl tRNA synthetase family. The polypeptide can be expressed
CC	recombinantly by transforming host cells with vectors comprising the v
CC	gene. Vals polypeptides and polynucleotides are useful for diagnosing
CC	diseases due to an infection of an organism and also the stage and type
CC	of the infection. The polypeptides are also used for screening for the
CC	modulators which can be used in the treatment of diseases. Diseases
CC	diagnosed, prevented or treated include otitis media, conjunctivitis,
CC	pneumonia, bacteremia, sinusitis, pleural empyema, endocarditis and
CC	especially meningitis. Vals polypeptides, polynucleotides and their
CC	(ant)agonists can prevent adhesion of bacteria to matrix proteins, and
CC	are useful for use on wounds and body implants to prevent bacterial
CC	infection.
XX	
SO	Sequence 2652 BP; 733 A; 612 C; 621 G; 686 T; 0 other;
XX	
SA	Alignment Scores:
SA	Pred. No.: 0.0177 Length: 2652
SA	Score: 130.00 Matches: 106
SA	Percent Similarity: 32.81% Conservative: 62
SA	Best Local Similarity: 20.70% Mismatches: 166
SA	Query Match: 3.50% Indels: 178
SA	DB: Gaps: 25
XX	
US	US-10-008-355-2 (1-712) x AAX32814 (1-2652)
XX	
OY	192 AspyAlphe-----LysAspValAlrGmetValPheAlaPro 203
XX	
DB	76 GATGTTTTCAGACCTTCAGCGCATCAAAAGCGTAAACCTTATTCATCTTTTCCACCA 135
XX	
OY	204 ProSerSerValGlyLysPheGlyGlyAspThrAspAsnTrpMetTrpProArgHisThr 223
XX	
DB	136 CCAAAACGTTACAGGTAACCT-----CACCTT 162
XX	
OY	224 GlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgProAlaGluTyrSer 243
XX	
DB	163 GGT-----CACCT 171
XX	
OY	244 LysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetClnGlyTyrLys 263
XX	
DB	172 TGGGATACA--ACTTTGCAAGATATTATTCATTCGTCACAAAACGATGCGTTTAT 228
XX	
OY	264 AlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrIleu---Thr 282

```

Db 229 -----ACCTTTGGCTTCCTGGGATGGACACGACGAGGATTCGCACT 270
Qy 283 SerTrpGlyValGluSparGllEGluASnGluASnSProArgllEGluValArgGly 302
Db 271 CAGGCTAAGTAAAGGACCTTCGTGTGGAGGCAATTCCTCATGACCTTGTCGT 330
Qy 303 -----IleYsgInGlyIleTrpLysGluAlaMetSerAlaSPGlnAlaThr 318
Db 331 GAGCTTTCTTGGACGAAGAAGCTCGGAATGGAAAGAC-----GAAATGCGCACT 378
Qy 319 ArgIleLysTrpAla-----SerLysTrpAlaGlnSerAlaAsnTrpLysAsnSerIle 337
Db 379 ACTATCAAGGAACAATAGGGGCAAGATGGGCTCTCTGTACATATCTCTGTAGCGCTTC 438
Qy 338 GlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGluArg 357
Db 439 ACTCTTGACGACGAAGTTGTCA-----AAAGCTGTCGTAAAG 474
Qy 358 AlaPheAlaAspTrpIleArgLys-----AsnGlyLysSerAlaValArgGly 373
Db 475 GTCCTTTGTGGACCTTTACAAGAAAGGCTGGATCTACCGGTGTAGATTATCATCACTCG 534
Qy 374 Asp-----ValLeuSerSerLeuGluLysAlaTrpLysGluGlyAlaLys 388
Db 535 GACCCAGACAGCTCGACAGCCCTTTCTGATATAGAGGTATCAACAAGATGTGGAAAGT 594
Qy 389 AlaAsnArgLysMetThrTrpLeu----- 396
Db 595 GCCTTACACCATGAATTCATGCTGGAAAGATGTTACGCGTCTTGAAGTTGTCTACA 654
Qy 397 -----SerGluThrLeuPheGlyGlyThrGluValValArgPheAlaGlnPheAlaAsn 414
Db 655 ACTCGTCCTAGACCAATGTTGGGAGCTT----- 684
Qy 415 AlaLeuAlaThrAsnProAspAlaHisAlaGlyIleLeuLysSerLeuAspAspLysTyr 434
Db 685 GCGGTTCGCGTCAACCCAGAA-----GACCCGCGCTAC 717
Qy 435 LysAsp-----TyrLeuProSerLeuAspArgLysValLeuProAla 448
Db 718 AAGCACTTGATGTGTAATAATGTCATCTTCCA---ATCGCTAATAAACATCATCCCATC 774
Qy 449 MetLeuAsp----- 455
Db 775 GTTGGAGATGACGACGAGATCTGAGTTGGTACTGCTGCTGAAATACACACTGCC 834
Qy 456 ArgIleProAlaAspLysLeuProAspIlePheLysAsnValIleAspLysPheLys 475
Db 835 CACGATCCAAATGACTCTTGTGGCCACAGCATATACTTCCACACAAGCATCAGTCATG 894
Qy 476 GlyAspThrLysLysTrpAlaAspPheValPheAspLysSerValValProTyrSerAsp 495
Db 895 AAGGACACGCAACCATGAATGACTTGGCTTGAATTTCAAGGCATG-----GAC 945
Qy 496 LysPheHisAlaMetLeuLysSerMetAspLysGluLysPheAlaLysAlaIleGluLys 515
Db 946 CGTTTGAAGCTCGT-----AAGGCACTGCTTAACTTAAGTGAAGAAATC 990
Qy 516 AspProAlaValGluLeuSerLysSerValIleAlaAlaAlaAlaAlaAlaAlaAla 531
Db 991 GGGGCCCTGTCATAAATCGAAAAAAGCGTGCACAGTGTGTGTCATCAGACGCTACAGT 1050
Qy 532 -----IleGlnAla 534
Db 1051 GTTGTGTGTAACCTCGCTTGTACTCAATGGTTCGTCAAGATGACCAATTTGGCTAAG 1110
Qy 535 AspAlaMetAlaAsnAlaTrpAlaIleGluLysGlyLysArgLeuPhePheAlaGlyLeu 554
Db 1111 AAGCGCACTTCCACACGAACGACAGACAGACAG----- 1143
Qy 555 ArgGluMetLysTrpGlyArgAlaLeuProSer----- 565

```

```

Db 1144 GTCGAATTCACCCACACCTGTCATCAAGATACCTTCCTCATGTAGTGAATAATGTCAC 1203
Qy 566 AspAlaAsnPheThrMetAlaGlyMetSerTrpGlySerIleLysGlyTrpGluProGlnAsp 585
Db 1204 GACTGGATTATCTCTGTCGACGCTCTGTGGGTCAC-----CAAAATC 1245
Qy 586 GlyAlaTrpTrpAsnTrpHisThrTrpGlyLysGlyValLeuGluLysGlnAspProLys 605
Db 1246 CTGCTCCCTGATCAAT-----GCTGATGCTGAATATGATCTCGCCGCAAGACTCCAGAA 1299
Qy 606 SerAspLeuPheAlaValGlnGluAsnIleLeuAsp 617
Db 1300 GGTGACGATGACCTCAGACGAAGACGCTCTTGAT 1335
RESULT 6
AAAl4369
ID AAAl4369 standard; DNA; 2652 BP.
AC AAAl4369;
XX
XX 15-ANG-2000 (first entry)
DE DNA encoding Streptococcus pneumoniae valyl tRNA synthetase (vals).
XX Valyl tRNA synthetase; vals; inhibitor; stringent response;
XX drug screening; antibacterial; antibiotic; genetic immunisation;
XX antibody; bacterial infection; meningitis; ds.
XX Streptococcus pneumoniae strain 0100993.
XX
XX Key Location/Qualifiers
FH CDS 1..1652
FT /tag a
FT /product "Streptococcus pneumoniae valyl tRNA synthetase
FT (vals)"
XX
XX US6051413-A.
XX
XX 18-APR-2000.
XX
XX 24-SEP-1998; 980S-0159539.
XX
XX 18-APR-1996; 96GB-0007791.
XX 17-OCT-1997; 97US-0953492.
XX 18-APR-1997; 97US-0844064.
XX
XX (SMIK ) SMITHKLINE BEECHAM CORP.
XX
XX Brown JR, Lawlor EJ, Wang M, Jaworski DD;
XX WPI; 2000-338311/29.
XX P-PSDB; AAY90514.
XX
XX Novel VALS polypeptides of valyl tRNA synthetase family useful for
XX treating otitis media, conjunctivitis, pneumonia and bacteremia
XX comprises a specified amino acid sequence -
XX
XX Claim 13; Columns 7-10; 21pp; English.
XX
XX This sequence represents DNA encoding Streptococcus pneumoniae valyl tRNA
XX synthetase (vals; NCIMB deposit No. 40794). Streptococci are the cause of
XX several types of human diseases, including otitis media, conjunctivitis,
XX pneumonia, bacteremia, sinusitis, pleural empyema, endocarditis and
XX especially meningitis. The frequency of S. pneumoniae infections has
XX risen dramatically over the past 20 years, probably due to the emergence
XX of multiply antibiotic resistant strains and an increasing population of
XX immunocompromised people. Vals represents a target for new antibacterial
XX agents. Inhibition of tRNA synthetases such as vals leads to a reduction
XX in the levels of charged tRNA, which triggers a cascade of responses
XX (known as the stringent response) resulting in a state of dormancy in the
XX bacterium. Vals, its variants and fragments, anti-vals antibodies, vals
XX inhibitors and nucleotides encoding vals may be used in the diagnosis,
XX prevention and treatment of bacterial infections such as meningitis.

```

CC Vals can be used to screen compounds for inhibitory activity. Vals
 CC may also be useful as an antigen for vaccination of a host to produce
 CC specific antibodies which protect against bacterial invasion into
 CC damaged tissues. Such antibodies could, for example, prevent the
 CC adherence of bacteria to wounds. Nucleotides encoding vals may be used as
 CC diagnostic reagents and therapeutic or prophylactic agents, particularly
 CC for genetic immunisation.

XX
 SQ Sequence 2652 BP; 733 A; 612 C; 621 G; 686 T; 0 other;

Alignment Scores:

Pred. NO.:	Length:	Score:	Matches:
130.00	2652	130.00	106
Percent Similarity:		32.81%	Conservative: 62
Best Local Similarity:		20.70%	Mismatches: 166
Query Match:		3.50%	Indels: 178
DB:		21	Gaps: 25

US-10-008-355-2 (1-712) x AAAA14369 (1-2652)

```

Oy 192 AspValPhe-----LysAspValArgMetValPheAlaPro 203
    |||||
Db 76 GATGTTTCAAGCCTTCAGCGCATCAAAAGCCTTATCATATTCACCA 135
Oy 204 ProSerSerValGlyLysPheGlyLysPheThrAspAsnTrpMetTrpProArgHisThr 223
    |||||
Db 136 CCAACGTTACAGGTAACCT-----CACCTT 162
Oy 224 GlyAspPheSerValPheArgValTyralGlyAlaAspAsnArgProAlaGluTyrSer 243
    |||
Db 163 GGT-----CACGT 171
Oy 244 LysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGlnGlyTyrLys 263
    |||
Db 172 TGGGTTACA---ACTTTCAGAGATTTATCATCCCTCAAAACGATCGAAGTTTAT 228
Oy 264 AlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrLeu---Thr 282
    |||||
Db 229 -----ACCTTGGCTCTCGGATGACACACGACGAGGATGCCACT 270
Oy 283 SerTrpGlyValGluAspArgIleGluAsnGluAsnAsnProArgIleGluValAlaArgGly 302
    |||||
Db 271 CAGGCTAAGTATGAGGAGCGCTGCGTGGTGAGGCATTTCCCGCTAAGACCTTGCTGCT 330
Oy 303 -----IleLysGlnGlyIleTrpLysGluAlaMetSerAlaAspGlnAlaThr 318
    |||||
Db 331 GAGCTTCTTGACGAAGTCTGGGATGGAAGAC-----GAATATGCCACT 378
Oy 319 ArgIleLysTyrAla---SerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsnSerIle 337
    |||||
Db 379 ACTATCAAGGAACAATGGGGCAAGTGGGCTCTGTAGACTATTCGTGACGCTTTC 438
Oy 338 GlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGluGluArg 357
    |||||
Db 439 ACTCTTGACGAAGTTTGTCA-----AAAGCTTGTGTAAG 474
Oy 358 AlaPheAlaAspTrpIleArgLys-----AsnGlyLysSerAlaValTyrGly 373
    |||
Db 475 GTCCTTGTGACCTTTACAAGAAAGCTGATCTACCGGAGGACTTATATCAACAATCG 534
Oy 374 Asp-----ValLeuSerLeuGluLysAlaTyrLysGluGlyAlaLys 388
    |||
Db 535 GACCCAGACGCTCGACACAGCCCTTGTGATATGAGTGATTCACAAAGATGTGGAAGGT 594
Oy 389 AlaAsnArgGluMetThrTyrLeu----- 396
    |||
Db 595 GCCTTCTACACAGATTAATTTACATGCTGGAAGATGTTACACGCGCTTGAAGTCTCA 654
Oy 397 -----SerGluThrLeuPheGlyLysThrGluValAlaArgPheAlaGlnPheAlaAsn 414
    |||||
Db 655 ACTGCTCTGAGACCATGTTTGGGACGTT----- 684
Oy 415 AlaLeuAlaThrAsnProAspAlaHisAlaGlyIleLeuLysSerLeuAspAspLysTyr 434
  
```

```

Db 685 GCGGTGGGTCACCAACCAGAA-----GACCCGCGCTAC 717
Oy 435 LysAsp-----TyrLeuProSerLeuAspArgLysValLeuProAla 448
    |||||
Db 718 AAGGACTTGATGTGTAATAATGATCCTTCGA---ATCGTAATTAACCTATCCCAATC 774
Oy 449 MetLeuAsp-----IleValArgArg 455
    |||
Db 775 GTTGAGATGAGCAGCAGATCCTGATTTGGTACTGGTGTGCGTAATAATCACACTGCC 834
Oy 456 ArgIleProAlaAspLysLeuProAspIlePheLysAsnValIleAspLysLysPheLys 475
    |||
Db 835 CACGATCCAAATGACTTCTGTTGGCCCAACGATCACTTGCACCAAGTCACAGTCATG 894
Oy 476 GlyAspThrLysLysTyrAlaAspPheAlaPheAspLysSerValValProTyrSerAsp 495
    |||
Db 895 AACGACGACGGAACCAATGACTTGGCTTTGAATTTTCAGGCATG-----GAC 945
Oy 496 LysPheHisAlaMetLeuLysSerMetAspLysGluLysPheAlaLysAlaIleGluLys 515
    |||||
Db 946 CGTTTGAAGCTCGT-----AAGCAGTCGTTGCTAAGTTGGAAGAAATC 990
Oy 516 AspProAlaValGluLeuSerLysSerValIleAlaAlaAlaArgAla----- 531
    |||||
Db 991 GGTGCCCTCGTCAAAATGCAAAACGTGTCACAGTGTGGTCACTACAGACGTACAGAGT 1050
Oy 532 -----IleGlnAla 534
    |||
Db 1051 GTTGTGTTGAACCTCGCTTCTACTCAATGTTCTGCAAGATGACCAATTCGCTAAG 1110
Oy 535 AspAlaMetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPheAlaGlyLeu 554
    |||||
Db 1111 AACGCAATGCAACCAACCAAGACAGAGACAGAA----- 1143
Oy 555 ArgGluMetTyrProGlyArgAlaLeuProSer----- 565
    |||
Db 1144 GTCCGAATTCACCCACCCCTGTTCAACGATACCTCTCAATGATGAAATGTCAC 1203
Oy 566 AspAlaAsnPheThrMetArgMetSerTyrGlySerIleLysGlyTyrGluProGlnAsp 585
    |||
Db 1204 GACGCGTTATCTCTCGACACTCTGGTGGGTAC-----CAATC 1245
Oy 586 GlyAlaTrpTyrAsnTyrHisThrThrGlyLysGlyValLeuGluLysGlnAspProLys 605
    |||||
Db 1246 CTGCTGCTGTAACAT-----GCTGATGTGAATAATGATGTGCGGCAAGACCTCCAGAA 1299
Oy 606 SerAspGluPheAlaValGlnGluAsnIleLeuAsp 617
    |||
Db 1300 GGTGACGGATGACTCAGACGAAGACGTGTGAT 1335
  
```

RESULT 7
 AAV52160/c
 ID AAV52160 standard; DNA; 6004 BP.
 AC AAV52160;
 DT 23-OCT-1998 (first entry)
 XX Streptococcus pneumoniae genome fragment SPO ID NO:27.
 DE Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
 KW computer readable medium; vaccine; pharmaceutical; composis; ds.
 XX Streptococcus pneumoniae.
 OS
 XX
 XX W09818931-A2.
 FN
 XX
 XX 07-MAY-1998.
 PD
 XX 30-OCT-1997; 97WO-US19588.
 PF
 XX 31-OCT-1996; 96US-0029960.

XX (HUMA-) HUMAN GENOME SCI INC.
 XX PA
 XX Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M;
 PI Kunsch CA, Rosen CA;
 XI
 XX WPI: 1998-272225/24.
 DR
 XX
 PT Computer-readable medium with recorded streptococcus pneumoniae
 PT polynucleotide sequences - useful in diagnostic kits and assays, and
 PT pharmaceutical compositions and vaccines for Streptococcus
 PT pneumoniae
 PS
 XX Claim 1: Page 305-309; 1409pp; English.
 CC The present invention describes a computer readable medium which has
 CC the nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524)
 CC recorded on it, or a representative fragment or a sequence at least 95%
 CC identical to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in
 CC SEQ ID NO:1 to 391 (AAV52134 to AAV52524) are genomic fragments from
 CC Streptococcus pneumoniae. The present invention also describes an
 CC isolated nucleic acid molecule encoding a homologue of any of the
 CC fragments of the S.pneumoniae genome (SEQ ID NO:1 to 391) where the
 CC nucleic acid molecule is produced by a process comprising: (a) screening
 CC a genomic DNA library using as a probe a target sequence defined by any
 CC of the sequences in SEQ ID NO:1 to 391, identifying members of the
 CC library which contain sequences that hybridise to the target sequence and
 CC isolating the nucleic acid molecules from the members; or (b) isolating
 CC mRNA, DNA or cDNA produced from an organism, amplifying nucleic acid
 CC molecules whose nucleotide sequence is homologous to amplification
 CC primers derived from the fragment of the S. pneumoniae genome to prime
 CC the amplification and isolating the amplified sequences. The computer
 CC readable medium can be used in a computer-based system for identifying
 CC fragments of the S. pneumoniae genome of commercial importance, or
 CC expression modulating fragments of the S. pneumoniae genome. Products
 CC from the present invention can be used in diagnosis kits and assays, and
 CC pharmaceutical compositions and vaccines for S. pneumoniae.
 XX
 SO Sequence 6004 BP; 1733 A; 1296 C; 1115 G; 1660 T; 0 other:
 Alignment Scores:
 Pred. No.: 0.0661 Length: 6004
 Score: 129.00 Matches: 116
 Percent Similarity: 33.62% Conservative: 77
 Best Local Similarity: 20.21% Mismatches: 194
 Query Match: 3.47% Indels: 187
 Gaps: 26
 DB: 19
 US-10-008-355-2.(1-712) x AAV52160 (1-6004)
 QY 131 LysIIleValIysValIhrspIysValIgluGIleuLysGIlyIlethrAspGluMet 150
 DB 4315 AAAATATAATATAGCATTTGTCTGCAGAAAGCTGCTAAGCAAAATTTCAAAATAAATA 4256
 QY 151 GluArgLeu-ArgLysAlaGlnGluValIcysGlnGluLeuAlaLysLysGluSnaIAs 170
 DB 4255 AACAAATATAGAGAAAGACACACATGTCTAAAGAACTTCCATCAATACAAATCCAGC 4196
 QY 170 pGluAsnGlnLeuCysIleValIgluProPheTyrSerAsnAsnGluTyrPheLeuIleVal 190
 DB 4195 CGAGCTTGAAGCTGTGCTGCTACCAAAATGCG-----CTTGA 4160
 QY 190 LTYrAspValPhe-----LysAspValArgMetValPheal 202
 DB 4159 TGGCGAATGTTTCAACGCTTCAAGCGATCAAAAGGCTAAACCTTATTCATCGGTATCC 4100
 QY 202 apcProSerSerValGlyLysPheGlyGlyAspThrAspAsnTrpMetTrpProArgH 222
 DB 4099 ACCACCAACGTAACGTGGAAACT-----CA 4073
 QY 222 sthngIyAspPheSerValPheArgValItyrAlaGlyAlaAspAsnArgProAlaGluTyr 242
 DB 4072 CTTTGGT-----CA 4064

QY 242 rSerLysAspAsnLysProTyrLysProValItyrPheAlaAlaValSerMetGlnItyr 262
 DB 4063 CGCTTGGGATACG---ACCTTGCAGATATCATCCGCCAAACCAAGCATGCAAGCTT 4007
 QY 262 rLysAlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrLeu-- 281
 DB 4006 TGAT-----ACCTTGGCTTCCAGGATAGGACACCGCTGTATTCG 3965
 QY 282 -ThrSerTrpGlyValGluAspArgIleGluAsnGlnLysAsnAsnProArgIleGluValAr 301
 DB 3964 GACTCAGCGTAAAGTTGAGAGCGCTTACGCTGTGAGGATATTCACCTATGACCTAAG 3905
 QY 301 gGly-----IleLysGlnGlyIleTrpLysGluAlaMetSerAlaAspGlnAl 317
 DB 3904 TCCTGAGCTCTTTTACGAGAGGCTGCGAATGGAAGC-----GAAATATGC 3857
 QY 317 arthrArgIleLysTyrAla---SerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsnSe 336
 DB 3856 CACTACCATCAAGCAACAAATGGGCAAGATGGGCTTCTGTGACTATTCGTGAGCG 3797
 QY 336 rIleGlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGlu 356
 DB 3796 TTTCACTCTTGATGAAGGTTTGCA-----AAAGCTGTTCG 3761
 QY 356 uArgAlaPheAlaAspTrpIleArgLys-----AsnGlyLysSerAlaValty 372
 DB 3760 TAAAGCTTTTGACACCTTTTACAGAAAGGCTGGATCTTCGTGGAGATTATCATCA 3701
 QY 372 rGlyAsp-----ValLeuSerSerLeuGluLysAlaTyrLys----- 384
 DB 3700 CTGGGACCCAGCAGACGTCGCACACCCCTTCTGTATTTAGAGGTATTCACAAAGATGAG 3641
 QY 385 -----GluGlyAlaLysAlaAsnArgLume 393
 DB 3640 AGGTGCTTCTTACACATGATATTCATGCTGGAAGATGTTACGCGCCCTTGAAGTTGC 3581
 QY 393 tThrTyrLeuSerGluThrLeuPheGlyIlyThrGluValAlaArgPheAlaGlnPheal 413
 DB 3580 TACAACTGCTGCTGAGACTATGTTGGGACGT----- 3547
 QY 413 aAsnAlaLeuAlaThrAsnProAspAlaHisAlaGlyIleLeuLysSerLeuAspAsp 433
 DB 3546 ---GGGCTTGGCGGTTAATCCAGAA-----CACCCGCG 3518
 QY 433 sTyrLysAsp-----TyrLeuProSerLeuAspArgLysValLeuP 447
 DB 3517 CTACAAGACTTGATTTGTTAAAAACGTATCTTCCA---ATCGTAATTAACATCATCC 3461
 QY 447 oAlaMetLeuAsp-----IleValAr 454
 DB 3460 AATCGTTGGAGATGAGACAGCAGACGATCTGAGCTTGCTGCTGTAATAATTCACACC 3401
 QY 454 gatArgArgIleProAlaAspLysLeuProAspIlePheLysAsnValIleAspLysSph 474
 DB 3400 TGGCCACGATCCAAATGACCTTGTGGTGGCCAAAGCTGCAATCTTGCCACAATGCAAGT 3341
 QY 474 eLysGlyAspThrLysLysTyrAlaAspPheValPheAspLysSerValValProTyrSe 494
 DB 3340 CATGACGACGACGAGCAATGATGAGCTTGCTTGAATTTTCAGCGATG----- 3289
 QY 494 rAspLysPheHisAlaMetLeuLysSerMetAspLysGluLysPheAlaLysAlaIleG 514
 DB 3288 -GATCGTTTGAAGCTGT-----AAGGCAAGTCGTTGCTAAGTGAAGA 3245
 QY 514 uLysAspProAlaValGluLeuSerLysSerValIleAlaAlaAlaArgAla----- 531
 DB 3244 AATCGGTGCCCTTGCAAAATGCAAAACGTGTCACAGTGTGTCACCTCAACAAGCTAC 3185
 QY 532 -----IleG 533
 DB 3184 AGGTGCTGATGTTGAGCCACGCTTGTCTACTCAATGCTTTCGTAAGATGACCAATTTGC 3125


```

QY 415 A1A1eUa1aThrasProaSPa1aH1aG1y1leUeuYsSerLeuAspaSPlySTyr 434
    |||:||||| |||:|||||
Db 685 GCGGTGGGGTTATCCAGAA-----GACCCGCGGTAC 717
QY 435 LysAsp-----TyrLeuProSerLeuSPaRgLySValLeuProa1a 448
    |||:||||| |||:|||||
Db 718 AAGCACTTGATGTGTA AAAACGTCATCTTCCA---ATCGCTATTAACATCATCCCAATC 774
QY 449 MetLeuAsp-----I1eValArgArg 455
    |||:||||| |||:|||||
Db 775 GTTGGAGATGACGACGAGATCTGAGCTTGTGACTGCTGCGTGA AAATCCACACCTGCC 834
QY 456 Arg1leProa1aSPlySLeuProaSP1lePheLysaSVa11leAsPlySLeuPheLys 475
    |||:||||| |||:|||||
Db 835 CACATCCAAATGACTTCTTGTTGGCCACAGCATATACCTTGCCACAAGTCAACGTATG 894
QY 476 GlyAspThryLysTyrAlaAspPheValPheAspLysSerValValProTyrSerAsp 495
    |||:||||| |||:|||||
Db 895 AACGACGACGAACTATGATGATGAGCTTGCTTGATTTTCAGGCGATG-----GAT 945
QY 496 LysPheH1a1aMetLeuYsSerMetAspLysGluYsPheAlaLysAla11leGluLys 515
    |||:||||| |||:|||||
Db 946 CGTTTTCAGCTCGT-----AAGCAGCTCGTTGCTTAAGTTGGAAGAAATC 990
QY 516 AspProa1aValGluLeuSerLysSerVal11leAla1a1aArgAla-----531
    |||:||||| |||:|||||
Db 991 GGTGCCCTTGTCAAATGAA AAAACGTCATTCAGTCCGTCACTCAGAACGACAGGT 1050
QY 532 -----I1eGluAla 534
    |||:||||| |||:|||||
Db 1051 GTCGTAGTTGAGCCAGCTTGCTACTCATGTCATGTTGTCAGATGAGCAATTTGGCTAAG 1110
QY 535 Aspa1aMetAlaAsna1a1aTyrAla11leGluLysGlyLysArgLeuPhePheAlaGlyLeu 554
    |||:||||| |||:|||||
Db 1111 AACGCCATTTGGCAACAAGACACAGAGAGACAAG-----1143
QY 555 ArgGluMetTyrProGlyArgAlaLeuProSer-----565
    |||:||||| |||:|||||
Db 1144 GTCGAATTTCCACCACTCGTTTCAACGATACCTTCTCATATGATGGA AAATGTCAC 1203
QY 566 Aspa1aAsnPheThrMetArgMetSerTyrGlySerI1eYsGlyTyrGluProGluAsp 585
    |||:||||| |||:|||||
Db 1204 GACGTGGGTATCTCTCGTCAGCTTGGGGGTCAC-----CAATATC 1245
QY 586 Gly1a1aTrrPyrAsnTyrH1a1aThrGlyLysGlyValLeuGluLysGluAspProLys 605
    |||:||||| |||:|||||
Db 1246 CCTCCCTGGTGCATAT-----GCTGATGCTGAATGTATGCGCGAAGAGCTCCAGAA 1299
QY 606 SerAspGluPheAlaValGluAsnI1eLeuAsp 617
    |||:||||| |||:|||||
Db 1300 GGTGACGATGACTCAAGCAAGAAAGACTCTTGAT 1335

```

```

RESULT 9
ID AA231948 standard; DNA: 2142 BP.
XX AA231948;
XX
DT 26-JAN-2000 (first entry)
XX
DE M. catarrhalis strain LES1 tbpb gene.
XX
KW TbpB gene: Tbp2; transferrin binding protein; diagnosis; otitis media;
KW genetic immunisation; Moraxella infection; antigen; vaccine; detection;
KW antitumour antibody production; therapy; ss.
XX
OS Moraxella catarrhalis.
XX
PN MO9952947-A2.
XX
PD 21-OCT-1999.
XX
PF 12-APR-1999; 99MO-CA00307.

```

```

XX
PR 14-APR-1998; 98US-0059584.
XX
PA (CONN-) CONNADUHT LAB LTD.
XX
PI MYERS LE, Schryvers AB, Harkness RE, Loosmore SM, Du R, Yang Y;
PI Klein MH;
XX
DR WPI; 1999-620376/53.
XX
P-PSDB; AAY43380.
XX
PT Nucleic acid encoding transferrin binding protein 2 of Moraxella
PT catarrhalis, useful for diagnostics, immunization and recombinant
PT protein production
XX
PS Claim 2; Fig 6; 114pp; English.
XX
CC This sequence encodes the Moraxella catarrhalis strain LES1 transferrin
CC binding protein (Tbp2) of the invention. This sequence is also referred
CC to as the TbpB gene. The TbpB gene is used to produce recombinant Tbp2;
CC for identification or diagnosis of Moraxella, or for cloning related
CC species, using hybridisation assays; and for genetic immunisation against
CC Moraxella infections, e.g. otitis media. The Tbp2 proteins are useful as
CC antigens, either in vaccines (including components of conjugate vaccines
CC that contain antigens from other bacteria or from tumours, in which case
CC they elicit production of antitumour antibodies that may be coupled to
CC chemotherapeutic agents or biologically active agents) or to raise
CC antibodies (for use as diagnostic reagents and for treating Moraxella
CC infections), also for detecting Moraxella antibodies.
XX
SQ Sequence 2142 BP; 753 A; 458 C; 457 G; 474 T; 0 other;

```

```

Alignment Scores:
Pred. No.: 0.0561 Length: 2142
Score: 123.50 Matches: 125
Percent Similarity: 32.17% Conservative: 77
Best Local Similarity: 19.90% Mismatches: 273
Query Match: 3.32% Indels: 153
DB: Gaps: 25
US-10-008-355-2 (1-712) x AA231948 (1-2142)
QY 119 LeuPro1leProGlyLeuSerValLysTyrLeuArgLysI1eValLysValThrasPlyS 138
    |||:||||| |||:|||||
Db 466 ATTCCTTTTGAATAAAATCTTATGATTAACCTTA AAAAATCATCTCGAGTTGTAAGTAA 525
QY 139 ValGluGlyGluLeuLysGlyLysI1eThrasPgluMetGluArgLeuArgLysAlaGlu 158
    |||:||||| |||:|||||
Db 526 TTTCAGACCAAAAAGCGGATTTGAATAATACACAAGACTGACACAAAAGATTATCA 585
QY 159 ValCysGluGluLeuAlaLysGlyLysAlaAspLysGluAsnGluLeuCysI1eValGlu 178
    |||:||||| |||:|||||
Db 586 TCAGAGCAAAAAGCAAAAAGCTCAAGAGCGTTGAGACAATGCTTA-----633
QY 179 ProPheTyrSerAsnAsnGluTyrPheLeuI1eVal-----TyrAspValPheLys 195
    |||:||||| |||:|||||
Db 634 ACTCAATTTGCCCAAGAAAATATACAAGAGCTATTTGAGAAGCCCATGATTA AAAATCT 693
QY 196 AspValArgMetValPheAlaProProSerSerValGlyLysPheGlyGlyAspThrAsp 215
    |||:||||| |||:|||||
Db 694 GACGACAGC-----702
QY 216 AsnTrpMetTrpProAlaGlyH1a1aThrGlyAspPheSerValPheArgValTyrAlaGlyAla 235
    |||:||||| |||:|||||
Db 702 -----702
QY 236 AspaNaRgProAlaGluTyrSerLysAspaNulysProTyrLysProValTyrPheAla 255
    |||:||||| |||:|||||
Db 703 ---AACCGTATCTAGAAATATGCAAGTCTGTTTAACTATCTTCTGATATACCGCC 759
QY 256 Ala-----ValSerMetGluGly-----TyrLysAla 264
    |||:||||| |||:|||||
Db 760 ACCGACCAAGCAAAAACCAATTAATGCGTATATGATGCGTGTATATTAAGGC 819

```



```

QY 265 ASPSPFYALAMETHRIIEGLYPHEPROGLYSERTHRASPARGYLEUTHRSERTIP 284
    ::: ||| :::: ||| :::: |||
DB 820 ACGGAACCGCCAAAGAGTA-----CCACAACAAAGGCAAAATATAAGTTATTGG 873
QY 285 GLY-----VALGLUASPARGYIEGLUASNGLUASNASPROARGILEGLUVALARGLY 302
    ::: ||| :::: ||| :::: |||
DB 874 GACTTATGACATGCTCCACACTTGATTAACAAATACACG-----GATTGGCCAGCT 924
QY 303 ILE---LYSGINGLYIETRPYSGIUALMETSERIALASPOLN---ALATHRARGILE 320
    ||| :::: ||| :::: ||| :::: |||
DB 925 ATCGCCAGACAAACCCAGTGGCTTACTGTTCTACTGATGAGATATGCACGCTCTTG 984
QY 321 -----LYSTYRALASERLYSTYRALAGINSERIALASNTYR 332
    ||| :::: ||| :::: |||
DB 985 ACAGACAAATAATACAAAGCCAGTATACATGTGCATATGCTCATAGCATGAATTT 1044
QY 333 TRPLYSASNSERTILEGLY-----METASNARGLYLEUALAARG 345
    ||| :::: ||| :::: |||
DB 1045 GATGTTAATTTTCTGATTAATAAATAAAGCAAACTTATCGTATCATGTTATCAGGC 1104
QY 346 LEUASPVAILIEGLYARGYLSARGALAGLUARGALAPHEALASPTRIEARGLYS 365
    ||| :::: ||| :::: |||
DB 1105 ACAGCTGTACCCGCCAAAGCGCTTATAAATAGAACTGATATCCACGCCAACCGCTTC 1164
QY 366 ASNGLYLYSSEALAVALTyrGLYASPVALLEUSERTSERLEUGLUYSALATYRILYSGLU 385
    ||| :::: ||| :::: |||
DB 1165 CGTGGCAGTGCC-----ACCGCAAGCATTAACGA-----GAA 1197
QY 386 GLYALALYSALASNARGLUETHRTYR---LEUSERGLUTHREUPHEGLYLYTHR 404
    :::: ||| :::: |||
DB 1198 GACAGCAAAACCCACACCCCTTACCAGCATGCTACACAAAGCTGATGCTTTT 1257
QY 405 GLUVALVALARGPHEALAGINPHEALASNALALEUALATHRASPROASPLAHSIALA 424
    :::: ||| :::: |||
DB 1258 TATGACCAAAAGCGAGAGAGCTGCGAAGTAATTCTTAACCGATGACAAACTCTTT 1317
QY 425 GLYILELEULYSERLEUASPARGYLYSTYRILYSASPTYRLEUPROSERLEUSPARGLYS 444
    ||::: ||::: ||:::
DB 1318 GGGCTTTGGTGTAAAGAGATTAAGTAAAGAAAAACGAAGCCATCTAGATGCTAT 1377
QY 445 VALLEUPROALAMETLEUASPLEVALARGARGYLEPROALASPLYSEUPROASP 464
    |||
DB 1378 GCACCT-----GGG 1386
QY 465 ILEPHELYASNVALILEASP---LYSLYSPHELYSGLYASPTHRLYSYTRIALASP 483
    ||| |||
DB 1387 ACATTTAATATAACAATAAAGCAACACATTCACCCCTTTACCAAAACACATGCGAT 1446
QY 484 ---PHEVALPHEASPLYSESERVALVALPROTYRSEASPLYSPHEHISALAMETLEULYS 502
    ||| ||| :::: |||
DB 1447 AACTTTGGCAATGCAAAAAGTTGCTTGGTTCACGTCATTAATTGGTGCTAC 1506
QY 503 SERMETASPLYSGILULYSALALIEGLULYSASPROALAVAGLULEUSER 522
    ||| :::: |||
DB 1507 GATGCCACCAAAAGAAATTCACCAAAATTCACCAAGAC----- 1548
QY 523 LYSSEVALILALALALALARGALALIEGLINALASPALMETALASNALATYRALA 542
    ||| ||| :::: |||
DB 1549 AAGCCAACTTGTGCCAACAAGGCGGAGACT--TTGATGGTGAATGATGAAGTT 1605
QY 543 ILEGLULYS-----GLYLSARGLEUPHEALAGLYLEUARGLUETHYRPROGLY 560
    ||| ||| :::: |||
DB 1606 ATCGTCMAAAACGATGGCAAAAACCTTGATACCTAAATTTGGTCAGCTTACTGTCGT 1665
QY 561 ARGALA-----LEUPROSERASPALASNPHETHMETARGMETSERTYRGLY 576
    :::: |||
DB 1666 GATAGCATAGCGCTTTTACAAAGCGCAACCGCTACACACAGGCGAGAAAGCGCGTA 1725
QY 577 SERILELYSGLYTYRGLURPGINASPGYALATPTYASNTYRHISHR----- 593
    ||| ||| ||| |||
DB 1726 CCAACCACAGCAAAAGCAAAATATCTGGGAACTGGGTAGATACATCACAGAGCGGCG 1785

```

```

QY 594 THRGLYLYSGLYVALLEUGLUYSGLINASPROLYSSERASPLUPHEALAVAGLINGLU 613
    ||||| ||| |||||
DB 1786 ACAGCAAAAAGCTTATAATGAGGCCCAAGAT-----ATTGCTGAT 1824
QY 614 ASNILELEUASPLEUPHEARGHYRHYLSASNTRYGLYARGTYRALAGLUASNGLYGLINLEU 633
    :::: |||
DB 1825 TTTGACATTGACTTTGAGAGAAATACGTTAAAGCAAACTGACCAACCCAGCGGCACA 1884
QY 634 HISILEALAPHELEUSERASNSASPILETHRGLY----- 645
    ||| |||
DB 1885 GATCCCTGCTTTAATCAATCAAAAGCTGAATTCGACGCAATGCTGACAGCAAAAGCCAC 1944
QY 646 -----GLYASNSERGLYSERPROVAL 652
    :::: |||
DB 1945 ACCACCAAGCGAGCAGCAGGCTACAGATAGATTCTAGACAGTACAGCAAAATCATC 2004
QY 653 PHEASPLYASNGLYARGYLEULEGLYLEUALAPHEASPLYASNTRIPLUALAMETSER 672
    :::: |||
DB 2005 GTCATCGAAAATGCCGAATTTACTGGGCTTTTATGTCCAAATGCAAAAGAGATGGGC 2064
QY 673 GLYASPIIEGLUPHEGLUPROASP 680
    |||
DB 2065 GGGTCATTTACACAGCATACCGAT 2088

```

RESULT 10

AA164839
ID AA164839 standard; cDNA; 3271 BP.

AC AA164839;

DT 04-DEC-2001 (first entry)

DE XN-100 coding sequence.

XX XN-100

KW XRN-100; homologous recombination exchange and exoribonuclease protein; chromosomal disorder; chromosome breakage syndrome; cancer; inflammation; developmental disorder; immunological disease; haemopathy; gene therapy; ss.

OS Unidentified.

PN CN1303931-A.

PD 18-JUL-2001.

PF 22-OCT-1999; 99CN-0119819.

PR 22-OCT-1999; 99CN-0119819.

PA (SHAN-) SHANGHAI BORONG GENE DEV CO LTD.

PI Mao Y, Xie Y.

DR WPI: 2001-566159/64.

DR P-PSDB: AAG78650.

XX

PT New polypeptide-homologous recombination medium-chain exchange protein and exoribonuclease protein-XRN-100 for treating tumors, inflammation, immunological disease and blood disease -

PS Claim 6; Page 16-18(Disclosure); 27p; Chinese.

XX

CC The present invention provides the protein and coding sequences of homologous recombination medium-chain exchange and exoribonuclease

CC protein XRN-100. The sequences can be used in the treatment of chromosomal disorders, chromosome breakage syndrome, developmental

CC disorders, cancer, inflammation, immunological diseases and haemopathy. The present sequence is the coding sequence of the invention.

SQ Sequence 3271 BP; 983 A; 648 C; 752 G; 888 T; 0 other:

Alignment Scores: 0.239 Length: 3271

Score:	119.50	Matches:	1030
Percent Similarity:	34.768	Conservative:	67
Best Local Similarity:	21.068	Mismatches:	159
Query Match:	3.218	Indels:	161
DB:	22	Gaps:	28

US-10-008-355-2 (1-712) x AAI64839 (1-3271)

QY	217	TriphetProAaGhIstNGIyAspPheSerValPheArgValYAlaGlyAlaAsp	236
Db	1614	TGGATATATCCATTCTCATTTATGACCACATTGGCT-----TCAGACTTTGAAGGCATTGGCA	166
QY	237	AsnArgProAlaGluTyrSerLysAspAsnLysProTyrLysProValYAlaPheAla	256
Db	1668	GACATGCCATCTGATTTGGAGAAAGGTACGAAACCGTTAAACACATA-----	171
QY	257	ValSerMetGInGlyTyrLysAlaAspAspTyrAlaMetThrIleGlyPheProGlySer	276
Db	1716	-----GAAACAATTATGGGGGTA-----TTTCAGACGTGCA	174
QY	277	ThrAspArgTyrLeuThrSerTyrPglYValGluAspArgIleGluAsnGluAsnPro	296
Db	1746	AGTGGTAAATTTTCTCACTCCA-----	176
QY	297	ArgIleGluValArgGlyIleLysGInGlyIleTyrLysGluAlaMetSerAlaAspGln	316
Db	1767	-----TCATGGCGGAAGCATGAGTGCCTGAT	179
QY	317	AlaThrArgIleLysTyr	326
Db	1797	TCCTGATTAATTTGACCTTCATCTCGAAGATTTTGCATATGATTTGAATGGAGAAATAT	185
QY	327	AlaGlnSerAlaAsnTyrTyrLysAsnSerIleGlyMetAsnArgGlyLeuAlaArgLeu	346
Db	1857	GCA-----TGG-----CAGGTCTTCTCTCTTG	188
QY	347	AspValIleGlyArgLysArg-----AlaGluGluArgAlaPheAlaAspTyrIle	363
Db	1881	CCATTCGGATGATGACAGAGCGCTACGAGCTGCCCTTAGAAGATATCCACAGACTCAT	194
QY	364	---ArgLysAsnGlyLysSerAlaValTyrGlyAspAlaLeu-----	376
Db	1941	CCAGAGAAGACCAGAGAAGAACGCTTGAGAGGTGATCTTATTTGTGGGAAACATCAC	200
QY	377	---SerSerLeuGluYAlaTyrLysGInGlyLysGluYAlaLysAlaAsnArgLysMet	393
Db	2001	CCACTCCATGACTTCATTTAAGACCTGTACCAGACAGAGTTC-----	204
QY	394	ThrTyrLeuSerGluThrLeuPheGlyGlyThrGluValValArgPheAlaGlnPheAla	413
Db	2043	-----ACAGAGCCACTGGAG-----	205
QY	414	AsnAlaLeuAlaThrAsnProAspAlaHisAlaGlyIleLeuLysSerLeuAspArgLys	433
Db	2058	-----GTACCCCTGAAACTATGTCAATGGGAT-----CAAGGAAG	209
QY	434	TyrLysAspTyrLeuProSerLeuAspArgLys---ValLeuProAlaMetLeuAspIle	452
Db	2094	TTT-----TCTTGGATGAAGAAACCACTTCTTCCAGATCAAAATAGTATGT	213
QY	453	ValArgArgArgIleProAlaAspLysLeuProAspIlePheLysAsn---ValIleAsp	471
Db	2139	-----TCTCCGTGTCCATNTGTATGAAGGATGTGACACAGAAACATGTAGTCACT	218
QY	472	LysLysPheLysGluArgThrLysLysTyrAlaAspPheValPheAspLysSerValVal	491
Db	2187	ATTAAATTTTAA---GACCCACACACTTGCTGAAATATTAATTTTAAAGCTGTATATCTT	224
QY	492	ProTyrSerArgLysPheHisAlaMetLeuLysSerMetAspLysGluLysPheAlaLys	511
Db	2244	CCAGAGGACAAAGAACGACGACGACTGTGAAACCTGATGACAGTGGAAAAATCCAGCAAT	230
QY	512	AlaIleGluLysAspProAlaValGluLeuSerLysSerValIleAlaValaValArgAla	530

Db 2304 GGACGCGAGGGAAGCCTCAGCTTGCGTTTAAACCCTGGAC-----CGAGGGCT 2351
 Oy 532 IleglnalaaspalametaI-----AsnAlatryAlaIlegluysgLy-LysAr 548
 Db 2352 GTGCACTGTGATCGAGCAGCGCTTCAGGACTTTGGGCATGTGATGCCAAGAGCTCAAGA 24111
 Oy 548 gLeuPhephealaglyLeu-----ArgglumetyrProgl 560
 Db 2412 ACTGGCATTTTACAGCAATGCTGCACACACACACTGTGACTTACAGGGAACCTTATACAG 24711
 Oy 560 yArAlaLeuProserAspAlaAsnhehrlmelaugmetsertryrGlyserIleuysgl 580
 Db 2472 CCGGCTT-TTGAAGAGACACACCCAGATTTCCAAAACCTATGTA-----AA 2515
 Oy 580 yTYrGluPrGlnaspLyAlaTrpTyraAsnTyrlHstHrthdGlyLysglValleuGl 600
 Db 2516 TATAGAGCCCGCAGAT--TCCGG-----CGAGGTCCTCCCTCTTT 2557
 Oy 600 uLysglAspProLysSerAspGluPheAlaValGlnIuAsnIleLeuAsp----- 617
 Db 2558 CCACGACACAA--AGCTTGTACAGAGGCGTGGGCGCTGAACCTCGCTCCCATGAGACCG 2614
 Oy 618 -----LeuPheArghTrLysAsnTyrgLyArGtyrAlaGluAs 630
 Db 2615 GATGCTGCAAACCCAGATGCAGCCTTCCAGCCAAACAGTACAGAGTCACTGGGCC 2674
 Oy 630 nglYgluPheUhlleAlaPheLeuSerAsnAspIleHrthGlyLysAsnserGlyse 650
 Db 2675 TGGT-----GGTATTCACCCAGACAGCATGATCTGTGGAGGAGACAGAGGATA 2722
 Oy 650 rProvalPheaspLysangLyArG 658
 Db 2723 TCCC-----AGAGAAAGGAAG 2738
 RESULT 11
 AAV21209
 ID AAV21209 standard; DNA: 1664976 BP.
 XX AC
 XX AAV21209;
 XX AC
 Dt 10-NOV-1998 (first entry)
 XX DE
 DE Methanococcus jannaschii circular chromosome.
 XX KW
 KW Methanococcus jannaschii; methanogenic archaeon; circular chromosome;
 XX genome; autotrophic; extrachromosomal element; identification; ds.
 OS
 OS Methanococcus jannaschii.
 XX PN
 PN W09807830-A2.
 XX PD
 PD 26-FEB-1998.
 XX PF
 PF 22-AUG-1997; 97WO-US14900.
 XX PR
 PR 22-AUG-1996; 96US-0024428.
 PA (GENO-) INST GENOMIC RES.
 PA (UNIT) UNIT ILLINOIS FOUND.
 PA (UNJO) UNIV JOHNS HOPKINS SCHOOL MEDICINE.
 PI Bult CJ, Smith HO, Venter JC, White OR, Woese CR;
 DR WPI; 1998-169145/15.
 XX
 XX Complete genome sequence of methano-genic archaeon, Methanococcus
 PT jannaschii - useful in identification of M. jannaschii genome
 PT fragment
 XX
 XX Claim 13; Page 152-585; 614pp; English.
 CC The present sequence represents the complete 1.66-megabase pair genome

QY 629 -----GLUasnGlyGlnIleuHisIleAlaPheLeuSerAsnAspIle 643
DB 5929936 ATTATTAGAGCTACAGACAGCTGCTCTTACAGATATCAAGATTTAAAGATTAACATTAT 592995
QY 643 eThrGlyGlyAsnSerglySer-----ProValPheAspIlyAsnGlyArgLeuIleG1 661
DB 5929996 AGACGATCGCTAAACAGCTGATACACACGCTTGGAACAAAGATATATTAATG 593055
QY 661 yLeuAlaPheAspGlyAsnTrp 668
DB 593056 AGTTAAGTTCTGTGGAGACTGG 593077
RESULT 12
AAT42063/C
ID AAT42063 standard; DNA; 1830121 BP.
XX AAT42063;
XX AAT42063;
DT 14-SEP-1999 (first entry)
XX
XX Haemophilus influenzae complete genome sequence.
DE
XX
KM Genome; bacterium; Haemophilus influenzae; computer readable medium;
KW expression modulating fragment; regulation; gene expression; vector;
KW organism; open reading frame; ORF; ds.
XX
OS Haemophilus influenzae.
XX
XX WO9633276-A1.
XX
XX 24-OCT-1996.
PD 24-OCT-1996.
XX
XX 22-APR-1996; 96WO-US05320.
PF
XX 07-JUN-1995; 95US-0487429.
PR 21-APR-1995; 95US-0426787.
PR 07-JUN-1995; 95US-0476102.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA (UYO) UNITV JOHNS HOPKINS.
XX
XX Adams MD, Fleischmann RD, Smith HO, Venter JC, White O;
XX
XX WPI; 1996-485782/48.
DR
XX
XX Haemophilus influenzae Rd genome recorded on computer readable
PT medium - useful for identifying commercially important nucleic acid
PT fragments by homology searching
XX
XX Claim 1; Page 77.2-77.1091; 1291pp; English.
PS
XX This sequence represents the complete genome sequence of the bacterium
CC Haemophilus influenzae strain Rd. The invention relates to a computer
CC readable medium (CRM) having recorded upon it the complete H. influenzae
CC nucleotide sequence (I), a representative fragment of (I) or a nucleotide
CC sequence at least 99% identical to (I). By providing the full-length
CC genomic sequence in a computer readable form, it is possible to identify
CC commercially important nucleic acid fragments and expression modulating
CC fragments (EMFs) of the Haemophilus genome. The EMFs can be used to
CC regulate the expression of a nucleic acid molecule. Vectors and altered
CC organisms comprising the predicted ORFs can be used to produce any of the
CC polypeptide fragments of the H. influenzae Rd genome.
XX
SQ Sequence 1830121 BP; 567399 A; 350615 C; 347389 G; 564036 T; 682 other;

Alignment Scores:
Pred. No.: 2e+03 Length: 1830121
Score: 117.00 Matches: 147
Percent Similarity: 30.62% Conservative: 58
Best Local Similarity: 18.38% Mismatches: 206
Query Match: 3.15% Indels: 349
DB: 17 Gaps: 42

US-10-008-355-2 (1-712) x AAT42063 (1-1830121)
QY 170 AspGluAsnGlnLeuCys----- 175
DB 605223 GATGAAATTCATATTACTTACTGGGGTTAATATTTATCTCTTGTACAGAAATTAATA 605164
QY 176 -----IleValGluProPheTyrSerAsnAsnGlyTyrPheLeuIle 189
DB 605163 TCTTAATTAATAAATAATTTTCTCTCTTCTTTTAAAGATCCGCTATTTTCTT 605107
QY 190 ValTyrAspValPheIlyAspValArgMetValPheAlaProProSerSerValGlyLys 209
DB 605106 -----AGAAATAAGGAATATTTACTACACCTTAATAAG----- 605074
QY 210 PheGlyGlyAspThrAspAsnTrpMetTrpProArgHisThrGlyAspPheSerValPhe 229
DB 605073 -----GAAACCTCTATGATGAACAGACGCTGCTTTTATTCAAATTAATGTCGACC 605026
QY 230 ArgValTyrAlaGlyAlaAspAsnArgProAla----- 240
DB 605025 AGTATTTCTTGCAATTAGTCAAAACCGCTTTTGGCATGCAAAAGAAAGCATGTGAT 604966
QY 241 -----GluTyrSerLys 244
DB 604965 TTACCATTTGTGCAACAACAGATGTCATAGTTCTTGACCGATTTTGACTATTATATA 604906
QY 245 AspAsnLysProTyrLysProValTyrPheAlaIleValSerMet----- 259
DB 604905 GATGCAACCAACCGATTAATTCGTTTACTGTCGCGCAAGCCTTATTCGTCMAACACGT 604846
QY 259 ----- 259
DB 604845 GCTGAAGTAAAAAATAGCGTATTGTTAGTAAATGGGATTTATTAATCAAGTATCAATT 604786
QY 260 -----GlnGlyTyrLys 263
DB 604785 GCAGACTATCAAGCCGCAACAGCTATTAAGGGAATTAATCAACCTGCATGATTTGT 604726
QY 264 -----AlaAspAspTyrAlaMet---ThrIleGly----- 272
DB 604725 TTAAATGCGATGAATTAAGAACTGGTACTATTAGTACACAGAAATTAATGATGCTGC 604666
QY 273 -----PhePro----- 274
DB 604665 AATTATCTTGCAATGCCATCAACACAGCTAATATTCCTCAATTTGTAATTAACAACGTACTA 604606
QY 275 -----GlySerThrAspArgTyrLeuThrSerTrpGlyValGluAspArg-----Ile 290
DB 604605 AAGCGGGGACAGAAAGAACCTTATTCACACCTTATGTCATTCAGAAATAACACTGTTT 604546
QY 291 GLUasnGluAsnAsnProArgIleGluValArgGlyIleLysGlnGly----- 306
DB 604545 GATTAATCAAGGT-----AAACACATAAATTAATAAATGCTTACATCGGTTT 604498
QY 307 -----IleTrpLysGluAla---MetSerAlaAspGlnAlaThrArg 319
DB 604497 GTGCCACCGCAATTAATGATGGTTGGATTAAGCAAACTTCAAGCAAAAGTGAACCCGT 604438
QY 320 -----IleLysTyrAlaSerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsnSerIle 337
DB 604437 GATATTGTAAAAACGACACAAAATAATGTACTGAA----- 604402
QY 338 GlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGluGluArg 357
DB 604402 ----- 604402
QY 358 AlaPheAlaAspTrpIleArgTyrAsnGlySerAlaValTyrGlyAspValLeuSer 377
DB 604401 -----ATGAAGAAAAAAGG-----GCTGATATTTGTCTT 604372
QY 378 SerLeu-----GluLysAlaTyrLysGluGlyAlaLysAlaAsn--- 390
DB: 17 Indels: 349 Gaps: 42


```

Db 10986 TTTATTTCCTCGCTGAT-----TTGACACAGTTTCTAAACTCCCGGTAGAACTTGCA 11039
QY 284 -----TTPGLYVALGLUASPARGLIEGLUASGLUASNAASNPRAAGLIEGLUVALARG 301
Db 11040 CCCCACTGGCCCGTGG-----GTGACAAACCCAGAAACATGATAAG----- 11078
QY 302 GLYLIEGLSGLNGLYLIEFTPLYSGLUALAMETSERALASPLGLN-ALATHRARGILELY 321
Db 11079 -----TGGCCAGACCGGTGTGGTTCCTGACCTTGCCGCCCTGCACATAG 11120
QY 321 STYRALASERLYTYRALAGINSERIALASNTYRTPLYS----- 334
Db 11121 TATAGCCGCGCGTGCATCGGTGCGGTACATAGTGGGCCCTGCAGNTTTCGGGGACC 11180
QY 335 -----ASNSERILEGLMETASNAAGGLYLEUALARGLEUASPAV11IEGL 350
Db 11181 CTTGGGGTGTGTCTACTATCTCACAATAATTTGTCAAGGGCCAGGCTCAAAATCTTCCG 11240
QY 350 YARGLYSARG-----ALAGLUGLUARGALAPHEALASPTYRPLEARGLYSASNGI 367
Db 11241 GAGACAGTCTTCAACACCGGCGCAATTGAGTGAATTCGCGTGAATTCGATGACCGG 11300
QY 367 YLYSSEALAVALTyrGLYSPALLESERSELEUGLULYSALATYRGLYSLUGLYAL 387
Db 11301 GAGCGA-----GAAATGTGCTGAGTCCCTCCCTCCCATGCTTTCATGGCGACGT 11347
QY 387 ALYSALASNAARGLUMETHRTYRLEUSERGLUHPHLEUPHGLYGLYTHRGUVALVA 407
Db 11348 CAAAGGC-----ACTACCGTTGGAGAGTCAACCATGT 11380
QY 407 L-----ARPPHE-----ALAGLPHALASNALEUALATH 418
Db 11381 CACCTCCAAATACCTCCGCGCTTCCTCCCAAGAAATCAGTGGCGGTACTCGGGCTTC 11440
QY 418 TASNPROASPLAHISALAGLYLLEULYSSELEUASPLASTYR----- 434
Db 11441 AAGCCCGGGGAAAGCCCAAAACAGATTTGACATTTAAACAGATGTATCCACAGATCT 11500
QY 435 -LYSASPTYRLEU--PROSERLEUASPARGLYSVALLEUPROALAMETLEUASPLIEVA 453
Db 11501 CGAAGCTTACCTCCACCAAGAACCCAGTCCAAAGTGTGGAATAATGTGGACTTCAA 11560
QY 453 LARGARGARGILEPROALASPLYSLEUPROASPLIEPHLYSASNAVALLEASPLYSLY 473
Db 11561 GGAAGTTCAGCATGATGCTCGAAGAGCAGACAGGCGCTATTTCACCTTAAAGCGCCCA 11620
QY 473 SPHELYSGLY--ASPTHRLYSSTYRALAASPLPHEVALPHEASPLYSSEVALVALPR 492
Db 11621 TTTACACCTGTACACAGCTTCAAGATGATGCTGCTGATCATC-----CGAGTTCC 11668
QY 492 CTYRSEASPLYPHEHISALAMETLEULYSSEMETASPLYSGLULYSFHEALAYSAL 512
Db 11669 TGTTAAC-----TCTAC 11680
QY 512 ALIEGLULYSAPPROALAVAL-----GLUEUSERLYSESERVALILEALALAAAR 530
Db 11681 GGTGATNTTGGACCCCGCATGGGGCCCTTGCCCTTTCGAAGAAGAGTTTCGGGTCCAC 11740
QY 530 GALALIEGLALASPLALAMETALASNAVALTYRALALIEGLULYSGLYFASARGLEUPH 550
Db 11741 TCAATGGGAGCTGACCTCCAGTCACTCACTTAT-----GATTACGCTCCCAAAATCAT 11794
QY 550 EPHEALAGLYLEU-----ARGGLUMETTYRPROGLYARGALALEUPROSERASPLAAS 568
Db 11795 CCTGTCTAGTGCATACCAGTGTGAATAGCCCGCTGGGTACAAATCTGCGCGTGGGGA 11854
QY 568 NPHETHRMET-----ARGMETSERTYRGLYSERILELYSGLYTYRGLIUPROGLINASPGI 586
Db 11855 GTTCTCCCTTGAGAGATCCATGAGTGAAGTCAAAACACACTGGGGGTTCGAA---TGGATAC 11911
QY 586 YALATPTYRASNTYRTHISHTHRGLYLYSGLYVALLEUGLULYSGLINASPPROLYSSE 606
Db 11912 AGCG---TATCTGTACAGATTCAACCGGAACGCT-----GAGACAGTGGAGAGATTCAA 11962

```

```

QY 606 RASPLUPHEALAVALTGLUGLUSNILEUASPLUPHEARGTHRLYSASNTYRGLYAR 626
Db 11963 TGATGCGTTTCGCGCGCCAG-----AAAGGAAATTTATTAAGC 12004
QY 626 GTYRALAGLUASNGLYGLINLEUHSILE-----ALAPH 637
Db 12005 CACTGGCCACACGATGAGTTCATTTTCCCGGGCCCTGTCATTTGAAACCACTTAGG 12064
QY 637 ELEUSERASNASNPILIEHTHRGLYASNSERGILYSEPRVALPHEASPLYSASNGI 657
Db 12065 CTGAATTTGAATGAATGATG-----GGGTCCATGCAAAAGCCTCTTGACAAAATTTGG 12115
QY 657 YARGLEULIEGLYLEUALAPHE 664
Db 12116 CCAACTTTTGTGATGATCTTTC 12137

RESULT 15
AAC77776
ID AAC77776 standard; cDNA, 3598 BP.
XX
AC AAC77776;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human cancer associated gene sequence SEQ ID NO:170.
XX
KW Human; cancer associated gene; cancer antigen; detection; cancer;
KW diagnosis; cytostatic; proliferative; vulnereary; immunomodulator;
KW antidiabetic; antisthmatic; antirheumatic; antiarthritic; antiviral;
KW antiinflammatory; antithyroid; antiallergic; antibacterial; cardiac;
KW dermatological; neuroprotective; thrombolytic; coagulant; noctropic;
KW vasotropic; antiporiatic; antidiabetic; gene therapy; inflammation;
KW immune disorder; haematopoietic cell disorder; autoimmune disorder;
KW allergic reaction; graft versus host disease; organ rejection;
KW haemostatic; thrombolytic; cardiovascular disorder; infection;
KW neurological disease; drug screening; ss.
XX
OS Homo sapiens.
XX
PN WO20005350-A1.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000MO-US05882.
XX
PR 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2000-587533/55.
XX
P-PSDB; AAB43567.
XX
PT Novel isolated nucleic acids comprising sequences encoding peptides
PT useful for treating or diagnosing e.g. cancer -
XX
PS Claim 1; Page 748-749; 2352pp; English.
XX
ACAC77607 to AAC78448 encode the human cancer associated proteins given
CC in AAB4398 to AAB44239. The proteins can have activities based on the
CC tissues and cells the genes are expressed in. Example of activities
CC include: cytostatic; proliferative; vulnereary; immunomodulator;
CC antidiabetic; antisthmatic; antirheumatic; antiarthritic;
CC antiinflammatory; antithyroid; antiallergic; antibacterial; antiviral;
CC dermatological; neuroprotective; cardiac; thrombolytic; coagulant;
CC noctropic; vasotropic; antiporiatic and antidiabetic. The
CC polynucleotides and polypeptides can be used for preventing, treating or
CC ameliorating medical conditions and diagnosing pathological conditions.
CC Polynucleotides, polypeptides, antibodies, agonists and antagonists from
CC the present invention may be used to treat immune disorders by activating
CC or inhibiting the proliferation, differentiation or mobilisation of

```


CC immune cells, to treat disorders of haematopoietic cells, autoimmune
 CC disorders, allergic reactions, graft versus host disease and organ
 CC rejection, modulate haemostatic or thrombolytic activity, modulate
 CC inflammation, cancers, cardiovascular disorders, neurological disease and
 CC bacterial or viral infections. The peptides, nucleotides, antibodies,
 CC agonists and antagonists may be also be used in drug screens. AAC78449 to
 CC AAC78457 and AAB44240 represent sequences used in the exemplification of
 CC the present invention.

XX Sequence 3598 BP; 1135 A; 636 C; 793 G; 1020 T; 14 other;

Alignment Scores:

Pred. No.:	0.59	Length:	3598
Score:	116.00	Matches:	110
Percent Similarity:	33.80%	Conservative:	85
Best Local Similarity:	19.06%	Mismatches:	198
Query Match:	3.12%	Indels:	185
DB:	21	Gaps:	30

US-10-008-355-2 (1-712) x AAC77776 (1-3598)

```

OY 90 CysGLYTYrGLYAlAlleGInserGInserThValAspHisprTYrLeuArgAspGLY 109
   |||||:::||||
DB 39 TGTGTTTGGGTGAG-----CTCTCAGACCGA 65

OY 110 PheValSerArgThrMetGLYGLuGLuLeu-----Pro 120
   |||||:::||||
DB 66 TTTCTAGCTCCGCGCGGACAGGTGCAGAGTCGRCGTGTCGACATGCGCGCCT 125

OY 121 IleProGLYLeuSerValTYrTYrLeuArgLYsIleValLYsValThraspLYsValGLU 140
   |||||:::||||
DB 126 CCACCGCGC-GCCGGAGAGAGCGGATCCCAAGTGGCCCAAGTGAAGAAACAAACCCCGC 184

OY 141 GLYGLuLeuLYsGLYIleThraspGLuMetGLUArgLeuArgLYsAlaGLInGLuValLYs 160
   |||||:::||||
DB 185 GCTGAGGTACAG--ATACTGTCTGA--CAACTCTTAAGAGAGGCTTAAGAAAGAGAA 238

OY 161 GLInGLuLeu-----AlaLYsLYsGLUAsnAlaAspGLUAsnGLInLeuCYsIle 176
   |||||:::||||
DB 239 CTTCGACCTCTCTCCACCTCCACCTCAACAGAGATCACAGATGAAGAAATTA----- 292

OY 177 ValGLUProPheTYrSerAsnAsnGLUTyrPheLeuIleValTYrAspValPheLYsAsp 196
   |||||:::||||
DB 293 -----AATGATATTAATACTAAGAGAAAGAGAAAGACTTTTGAAGAT 331

OY 197 ValArgMetValPheAlaProProSerSerValGLYLYsPheGLYGLYAspThraspAsn 216
   |||||:::||||
DB 332 -----AATATAAGAAAAAAGACGACTGTGATTAGTAAC 364

OY 217 TrpMet-----TrpProArgHisThrGLYAspPheSerValPheArgValTYr 232
   |||||:::||||
DB 365 TGGATTAATAATACGCACAATG----- 385

OY 233 AlaGLYAlaAspAsnArgProAlaGLUTyrSerLYsAspAsnLYsProTYrLYsProVal 252
   |||||:::||||
DB 386 -----GAGAAAGAGCTTAAGAGAGATTTCAAGGCTTGATGCCA 424

OY 253 TYrPheAlaAlaValSerMetGLInGLYTYrLYsAlaAspTYrAlaMetThrIleGLY 272
   |||||:::||||
DB 425 TACGAGCGGTCTTAGATGA-----GACTAC----- 451

OY 273 PheProGLYSerThraspArgTYrLeuThrSerTYrPGLYValGLUAspArgIleGLUAsn 292
   |||||:::||||
DB 452 -----CGAATATATTACACTGTGCTGAATTAAGCAAGAAATGGAATG 493

OY 293 GLUAsnAsnProArgIleGLUValArgLYsIleLYsGLInGLYIleTrpLYsGLUAlaMet 312
   |||||:::||||
DB 494 AAGAAATCGCAAGTCAACCATGCTCGA-----AATATCTGGAGCGGCGCAAT 541

OY 313 SerAlaAspGLInAlaThrArgIleLYsTYrAlaSerLYsTYrAlaGLInSerLYsAsnTYr 332
   |||||:::||||
DB 542 -----ACAAGCTGCTCGAGTGAATTAATCAAGTTCTGTGATACAAGTACAGTAC 586

```

```

OY 333 TrpLYsAsnSerIleGLYMetAsnArgGLYLeuAlaArgLeuAspValIleGLYArgLYs 352
   |||||:::||||
DB 587 ARGAGAGAAATGTGTGGA-----AACGTTGCCGTCGCCCG 622

OY 353 ArgAlaGLInGLUArgAlaPheAlaAspTYrIleArgLYsAsnGLYSerLYsValTYr 372
   |||||:::||||
DB 623 CAGGTGTTTGACCGC-----TGATGTAGGTGGACACCTGAGAGCAACCTGG 670

OY 373 GLYAspValLeuSerSerLeuGLUAsnAlaTYrLYsGLInGLYAlaLYsAlaAsnArgGLU 392
   |||||:::||||
DB 671 CACTCTCATATC--ACATTGAGCTGAGATCAAAAGAG-----GTGATCGGCGCC 718

OY 393 MetThrTYrLeuSerGLUThrLeuPheGLYGLYThrGLUVal-----ValArgPhe 409
   |||||:::||||
DB 719 CCGACCATTTATGACGAGAKTTCCTCGTCGACCTGATGTTAAGACTGATCAAGTAT 778

OY 410 AlaGLInPheAlaAsnAlaLeuAla----- 417
   |||||:::||||
DB 779 GCCCGCTTGAAGAAACATGCTTATTTGCCCATGCACGGAAGTGTAGAGAGACT 838

OY 418 -----ThraspProAspAlaHisAlaGLYIle--LeuLYsSerLeu 430
   |||||:::||||
DB 839 GTGGAATTCCTTGGAGATGAACATATGATGACACCTTTATGTCCTTGCCAAAGTTT 898

OY 431 AspAspLYsTYrLYsAspTYrLeuProSerLeuAspArgLYsValLeuProAlaMetLeu 450
   |||||:::||||
DB 899 GAAGAAATATCAAGAAAGACTT-----GAAAGGTACACAGTATTTACAAAGTATGCCCTG 952

OY 451 AspIleValArgArgIleProAlaAspLYsLeuProAspIlePheLYsAsn----- 468
   |||||:::||||
DB 953 GAC-----AGAAATTTCAAAACAAGATGCCCAAGAACTCTTTAAATATATATACC 1000

OY 469 ValIleAspLYsLYsPheLYsPheLYsPheLYsTYrAlaAspPheValPheAspLYs 488
   |||||:::||||
DB 1001 AACTTGTGAAGAAAGACTT--GCTATAGCGCGGATATGAAGATATCATGTTGTGCAAA 1057

OY 489 SerValValProTYrSerAspLYsPheHisAlaMetLeuLYsSerMetAspLYs--GLU 507
   |||||:::||||
DB 1058 CGGAGATTCAGATGACGAGAAAGAGTGAAGCGCAATCCACAAATATATGATCATGCTTT 1117

OY 508 LYsPheAlaLYsAlaIleGLULysAspProAlaValGLULeuSerLYsSerValIleAla 527
   |||||:::||||
DB 1118 GATTACTTGCCTGTGTGAAAGTGCAGCAGAACTGAA----- 1156

OY 528 AlaAlaArgAlaIleGLInAlaAspAlaMetAlaAsnAlaTYrAlaIleGLULysGLYs 547
   |||||:::||||
DB 1157 GCGGTGAGAGAGAGTATGAAGAGGCCATTTGCCATGTGCCACCATTCAGAGAGAAAGG 1216

OY 548 ArgLeuPhePheAlaGLYLeuArgGLUMetTYrProGLYArgAlaLeuProSerAspAla 567
   |||||:::||||
DB 1217 CAC-----TGGNAGCGCTACACTTAT 1237

OY 568 AsnPheThrMetArgMetSerTYrLYsSerIleLYsGLYTYrGLUProGLInAspGLYAla 587
   |||||:::||||
DB 1238 -----CTT 1240

OY 588 TrpTYrAsnTYrHisThrThrGLYsGLYValLeuGLInLYsGLInAspProLYsSerAsp 607
   |||||:::||||
DB 1241 TGGATCAACTATGCACTCTATGAAGAA-----TTGAGAGCAAGAGATCTGAGAGGACA 1294

OY 608 GLUpheAlaValGLInGLUAsnIleLeuAspLeuPheArgThrLYsAsnTYr 624
   |||||:::||||
DB 1295 AGACAGGTGTATCAAGCTCT--TTGGAAGTAATTCCTCACAAAAGTTTC 1342

```

Search completed: October 17, 2002, 23:17:53
 Job time : 1848 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 17, 2002, 22:44:46 ; Search time 54 Seconds
(without alignments)
3238.723 Million cell updates/sec

Title: US-10-008-355-2

Perfect score: 3719

Sequence: 1 NOKAKISILGALLGASG.....LFMDKMGCCPRLIOELKII 712

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame_p2n.model -DEV=xlh
-Q=/cg2_1/USPTO_spool/US10008355/runat_16102002_121154_27717/app_query.fasta_1.903
-DB=Issued_Patents_NA -QFMT=fastp -SUFFIX=tnl -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blsum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10008355 -CGEN=1_1_18 -etrunat_16102002_121154_27717 -NCPV=6 -ICPU=3
-NO_XLPHY -NO_MAP -LARGEOUTER -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued_Patents_NA:*
1: /cg2_6/pdata/2/lna/5A.COMB.seq:*
2: /cg2_6/pdata/2/lna/5B.COMB.seq:*
3: /cg2_6/pdata/2/lna/6A.COMB.seq:*
4: /cg2_6/pdata/2/lna/6B.COMB.seq:*
5: /cg2_6/pdata/2/lna/PCTUS.COMB.seq:*
6: /cg2_6/pdata/2/lna/Backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	130	3.5	2652	2	US-08-953-492-1
2	115	3.1	4853	2	US-08-793-824-1
3	111	3.0	441529	4	US-09-103-840A-1
4	110	3.0	3279	4	US-08-446-137B-1
5	109	2.9	4403765	4	US-09-103-840A-2
6	106	2.9	2472	3	US-08-335-844A-7
7	106	2.9	3358	3	US-08-335-844A-20
8	106	2.9	3688	4	US-09-327-536-1
9	105.5	2.8	1994	3	US-08-600-982-22
10	105.5	2.8	1994	5	PCT-US94-10261A-22
11	104	2.8	3731	4	US-08-811-583-1
12	102.5	2.8	6142	4	US-09-514-302-1

13	102	2.7	2582	2	US-08-816-105A-2
14	102	2.7	4310	3	US-09-008-172-1
15	102	2.7	4310	4	US-09-210-361-5
16	101.5	2.7	2253	1	US-07-854-596B-39
17	101.5	2.7	3727	1	US-08-249-380-1
18	100.5	2.7	3278	1	US-08-484-105-13
19	100.5	2.7	3278	1	US-08-484-106-13
20	100.5	2.7	8501	1	US-09-298-367B-6
21	100	2.7	4330	3	US-09-310-293-1
22	100	2.7	4330	4	US-09-579-376-1
23	100	2.7	6744	1	US-08-119-125A-2
24	100	2.7	46899	1	US-08-471-119A-1
25	99.5	2.7	2167	3	US-08-861-083-83
26	99	2.7	3531	2	US-08-714-402-1
27	99	2.7	5433	3	US-09-157-021-35
28	99	2.7	5433	3	US-09-156-842-35
29	98.5	2.6	2012	4	US-08-026-143B-2
30	98.5	2.6	2012	4	PCT-US92-10621-2
31	98.5	2.6	2012	5	PCT-US94-02233-2
32	98.5	2.6	2217	1	US-08-712-241-4
33	98.5	2.6	5175	2	US-08-843-530B-3
34	98	2.6	2652	4	US-09-115-704-1
35	98	2.6	7253	4	US-09-268-347-35
36	98	2.6	7291	4	US-08-913-942-3
37	97.5	2.6	1404	3	US-08-961-083-9
38	96.5	2.6	1575	4	US-09-230-388-2
39	96.5	2.6	2136	2	US-08-949-941B-1
40	96.5	2.6	3728	1	US-08-111-939-1
41	96.5	2.6	4821	1	US-08-053-614-3
42	96.5	2.6	4821	1	US-08-316-397B-3
43	96.5	2.6	4821	2	US-09-034-306-3
44	96.5	2.6	4821	3	US-09-259-437-3
45	96.5	2.6	4821	5	PCT-US93-09782-3

ALIGNMENTS

RESULT 1
US-08-953-492-1
; Sequence 1, Application US/08953492
; Patent No. 5849555
; GENERAL INFORMATION:
; APPLICANT: Brown, James
; APPLICANT: Jaworski, Deborah
; APPLICANT: Lawlor, Elizabeth
; APPLICANT: Wang, Min
; TITLE OF INVENTION: NOVEL VALS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/953,492
; FILING DATE: 17-OCT-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/844,064
; FILING DATE: 18-APR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 9607991.8
; FILING DATE: 18-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Gilmul, Edward R
; REGISTRATION NUMBER: 38,891

Sequence 2, Appl1
Sequence 1, Appl1
Sequence 5, Appl1
Sequence 39, Appl1
Sequence 1, Appl1
Sequence 13, Appl1
Sequence 13, Appl1
Sequence 6, Appl1
Sequence 1, Appl1
Sequence 2, Appl1
Sequence 1, Appl1
Sequence 83, Appl1
Sequence 1, Appl1
Sequence 35, Appl1
Sequence 35, Appl1
Sequence 2, Appl1
Sequence 2, Appl1
Sequence 4, Appl1
Sequence 3, Appl1
Sequence 3, Appl1
Sequence 35, Appl1
Sequence 9, Appl1
Sequence 2, Appl1
Sequence 1, Appl1
Sequence 1, Appl1
Sequence 3, Appl1
Sequence 3, Appl1
Sequence 3, Appl1
Sequence 3, Appl1

```

: REFERENCE/DOCKET NUMBER: P31458-4/1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 610-270-4478
: TELEFAX: 610-270-5090
: TELEX:
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2652 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: Genomic DNA
: US-08-953-492-1

Alignment Scores:
Pred. No.: 0.00287 Length: 2652
Score: 130.00 Matches: 106
Percent Similarity: 32.81% Conservative: 62
Best Local Similarity: 20.70% Mismatches: 166
Query Match: 3.50% Indels: 178
DB: Gaps: 25

US-10-008-355-2 (1-712) x US-08-953-492-1 (1-2652)

OY 132 AspValPhe-----LysAspValArgMetValPheAlaPro 203
DB 76 GATGTTTCAAGCCTTACAGCGCATCAAAAGCGTATTATTCATCGTATTCACCA 135
OY 204 ProSerSerValGlyLysPheGlyLysAspThrAspAsnTrpMetTrpProArgHisThr 223
DB 136 CCAAAAGTTACAGGTTAACTT-----CACCTT 162
OY 224 GlyAspPheSerValPheArgValTyraGlyAlaAspAsnArgProAlaGluTrpSer 243
DB 163 GGT-----CACCT 171
OY 244 LysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGlnGlyTyrLys 263
DB 172 TGGGATTACA--ACTTTCAGATGATATTCATCCGCAAAAGCATGCAAGTTTGAT 228
OY 264 AlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrLeu--Thr 282
DB 229 -----ACCTTTGGCTTCTCGGGATGACCGACCGAGGATGCCACT 270
OY 283 SerTrpGlyValGluAspArgIleGluAsnGluAsnProArgIleGluValArgGly 302
DB 271 CAGCGTAGAGTAGAGAGCGCTTGGGTGAGGCATTTCCCGTAGACCTTGTCGT 330
OY 303 -----IleLysGlnGlyIleTrpLysGluAlaMetSerAlaAspGlnAlaThr 318
DB 331 GAGCTTCTTGACGAAAGCTCTGGAAATGGAAAGAC-----GAAATATGCCACT 378
OY 319 ArgIleLysTyrAla--SerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsnSerIle 337
DB 379 ACTATCAAGACATGAGGCGCAAGATGGGCGCTCTGTAGACTATCTCGTAGCGTTTC 438
OY 338 GlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGluGluArg 357
DB 439 ACTCTTGACGAGGTTTGCA-----AAAGCTGTCGTAG 474
OY 358 AlaPheAlaAspTrpIleArgLys-----AsnGlyLysSerAlaValTyrGly 373
DB 475 GTCCTTGAGACCTTACAAAGAAAGCTGGATCTACCGCTGGAGTTTATATCTCAACTGG 534
OY 374 Asp-----ValLeuSerSerLeuGluLysAlaTyrLysGlnGlyAlaLys 388
DB 535 GACCCAGACAGCTCGACAGCCCTTCTGTGATATGAGTGATTCACAAAGATGTGAAGT 594
OY 389 AlaAsnArgGluMetThrTyrLeu----- 396
DB 595 GCCCTTACACACATGATATACATGCTGGAAGATGTTACAGCGCTTCGAAGTTCCTACA 654
OY 397 -----SerGluThrLeuPheGlyLysThrGluValValArgPheAlaGlnPheAlaAsn 414

```

```

DB 655 ACTGCTCCTGAGACCATGTTGGGACGTT----- 684
OY 415 AlaLeuAlaThrAsnProAspAlaHisAlaGlyIleLeuLysSerLeuAspLysTyr 434
DB 685 GCGGTGGCGTCAACCCAGAA-----GACCCGGCTTAC 717
OY 435 LysAsp-----TyrLeuProSerLeuAspArgLysValLeuProAla 448
DB 718 AAGGACTTGATGTGTAATAATGTCATCTTCCA--ATCGCTAATTAACATCCCAATC 774
OY 449 MetLeuAsp-----IleValArgArg 455
DB 775 GTTGGAGATGAGCAGCAGATCTGAGTTGGTACTGGTGTGTAATAATCACAACCTGCC 834
OY 456 ArgIleProAlaAspLysLeuProAspIlePheLysAsnValIleAspLysPheLys 475
DB 835 CACGATCCCAATGACTTGTGGTGGCCACAGCTCATTAACCTTGCCCAAGTCACGTCATG 894
OY 476 GlyAspThrLysLysTyrAlaAspPheValPheAspLysSerValValProTyrSerAsp 495
DB 895 AACGACGACGAGACCATGATGACTTGGCTTGAATTTTCAGGCATC-----GAC 945
OY 486 LysPheHisAlaMetLeuLysSerMetAspLysGluLysPheAlaLysAlaIleGluLys 515
DB 946 CGTTTGAAGCTCGT-----AAGCGATCGTTGCTTAACCTTGGAAGAAATTC 990
OY 516 AspProAlaValGluLeuSerLysSerValIleAlaAlaAlaArgAla----- 531
DB 991 GGTGCCCTTCGTCAAAATGCAAAACGTGTCCACAGTGTGGTCACTCAGACGTCACAGT 1050
OY 532 -----IleGlnAla 534
DB 1051 GTTGTGTTGAACCTCGCTGTCTACTCATATGTTCTGTCAAGATGACCAATTTGGCTAAG 1110
OY 535 AspAlaMetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPhePheAlaGlyLeu 554
DB 1111 AACGCCATTTGCCAACCAACAGACACAGACAGACAG-- 1143
OY 555 ArgGluMetTyrProGlyArgAlaLeuProSer----- 565
DB 1144 GTCGATTCCTACCCACCTCGCTTTCACAGATACCTTCCTCAATGATGATAAATGTCAC 1203
OY 566 AspAlaAsnPheThrMetArgMetSerTyrGlySerIleLysGlyTyrGluProGlnAsp 585
DB 1204 GACGCGTATTCCTCGACAGCTCTGCGGGCTAC-----CAAAATC 1245
OY 586 GlyAlaTrpTyrAsnTyrHisThrThrGlyLysGlyValLeuGluLysGlnAspProLys 605
DB 1246 CCGCGCTGTACAAAT-----GCTGATGTTGAATGTATGTGGCGCAAGAAAGCTCCAGAA 1299
OY 606 SerAspGluPheAlaValAlaGlnGluAsnIleLeuAsp 617
DB 1300 GGTGACGATGAGACTCAGACGAAAGACTCTTGGAT 1335

RESULT 2
US-08-793-824-1
: Sequence 1, Application US/08793824
: Patent No. 5981838
: GENERAL INFORMATION:
: APPLICANT: Simpson, Christine Lynn
: APPLICANT: Giffard, Phillip Morrison
: APPLICANT: Jacques, Nicholas Anthony
: TITLE OF INVENTION: Genetic Manipulation of Plants to
: TITLE OF INVENTION: Increase stored Carbohydrates
: NUMBER OF SEQUENCES: 2
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Griffith Hack & Co
: STREET: Level 8, 168 Walker Street
: CITY: No. 5981838th Sydney
: STATE: New South Wales
: COUNTRY: Australia
: ZIP: 2060

```

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/793,824
FILING DATE:
CLASSIFICATION: 800
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: AU PM7643
FILING DATE: 24-AUG-1994
TELECOMMUNICATION INFORMATION:
TELEPHONE: 61 2 9957 5944
TELEFAX: 61 2 957 6288
TELEX: 26347
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4853 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHEetical: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Streptococcus salivarius
US-08-793-824-1

Alignment Scores:
Pred. No.: 0.0328 Length: 4853
Score: 115.00 Matches: 161
Percent Similarity: 34.10% Conservative: 106
Best Local Similarity: 20.56% Mismatches: 262
Query Match: 3.09% Indels: 257
DB: Gaps: 40

US-10-008-355-2 (1-712) x US-08-793-824-1 (1-4853)
QY 70 GlyGlyGlyCysThrGlyLeuThrValSerAsp-----GlnGlyLeuIle 84
   ||||| ||| :|||:||||| ||||| |||
DB 1776 GGTGGACATATGATGCGGATCTCTTGCGAGATGCTTGTACTTCAAGAAAAA 1835
QY 85 PheThrAsnHis-----HisCysGlyTyr-GlyAlaIleGlnSerGlnse 99
   ||||| ||| :|||:||||| ||||| |||
DB 1836 CCGTGTCCGAGATATGAAGCAAAATGCCATTCCTCATTTGCAATCTTGAAGCTTGTC 1895
QY 99 ThrValAspHisAspTyrLeuArgAspGlyPheValSerArgThrMetGlyGlnGluLe 119
   ||||| ||| :|||:||||| ||||| |||
DB 1896 ATACAAAGCAAGCAACCAATACAAAGAGAT-----ACCAAGGTGCTCAGTT 1940
QY 119 uProIleProGly-----LeuSerValIleTyrLeuArgLys----- 131
   ||| ||| ||| :|||:||||| ||||| |||
DB 1941 GTCTATCGCAATCCACTACGTGAAAGCTTTTGACGACTTCTTCCGTAAAGCAATTA 2000
QY 131 ----- 131
DB 2001 CCGTGGACCTTGGAGCGCGTATTAACCTCTTAACAATCGTTCAAGTGACAGAA 2060
QY 132 -----IleValIleValIleThrAsp----- 137
   ||||| ||| :|||:||||| ||||| |||
DB 2061 ACACAGCGACGATGATGCCAATTATATTTCGTGCGAGCCCATGATGAGAGTGCAAC 2120
QY 138 -----LysValGlnGlyGlnLeuLysGlyIleThrAspG1 149
   ||||| ||| :|||:||||| ||||| |||
DB 2121 TGTTCGCTAATATCATCAGCAGCATTAATCCAAAGAGATGTTCACTTCAC 2180
QY 149 uMetGluArgLeuArgLysAlaGlnGluValCysGln---GluLeuAlaLys-----Ly 166
   ||||| ||| :|||:||||| ||||| |||
DB 2181 TATGGTAGGCTCAACACAGCGCTTCGAAATTCATCAATGCGGAGATCCGCAAGCGTGA 2240
QY 166 sGluAsnAlaAspGluAsnGlnLeuCysIleValGluProPheTyrSerAsnGlnIuTy 186
   ||||| ||| :|||:||||| ||||| |||

```

```

DB 2241 AAAGTACACCCAGTACATATCCAGCTGCTATGACCAAAATGTTGACCAACAGGATAG 2300
QY 186 rPheLeuIleValIleTyr-----AspValPheLysAspValArgMetValPheAlaPro 204
   ||||| ||| :|||:||||| ||||| |||
DB 2301 TATCACTGCTGTTACTACAGGGAGCCCTTTACACAGATGCGCAATACATGCTGAAAA 2360
QY 204 oSer----- 205
   |||
DB 2361 ATCACCCTACTATTAATGCCATGATGCCCTGCTCCGTGCTGCATCAAAATATGTACAG 2420
QY 206 -----SerVal 207
DB 2421 TGTCAAGACATGAAAGTTACTAAGCTTAATGCTATGAATCATGTATCTGTGGCTTA 2480
QY 207 LcIlyLysPheGlyGlyAspThrAspAsnTrpMetTrp-----ProArg 221
   ||||| ||| :|||:||||| ||||| |||
DB 2481 TGTGTAAAG---GGGCGAGAAAGAACCCACCA--CCTTGCTACTGCTGAACACGCCAACAG 2536
QY 221 gHisThrGlyAspPheSer---ValPheArg-ValTyrValGlyValAlaAspAsnArg---- 238
   ||||| ||| :|||:||||| ||||| |||
DB 2537 GAATGCTGCTCTTACAGCCAAACCGTCCAGATATGAGTTGGGAGCTTAATGATCTCTGG 2596
QY 239 -----ProAlaGluTyrSerLysAspAsnLysProTyrLysProValTyrPheAlaAla- 256
   ||||| ||| :|||:||||| ||||| |||
DB 2597 TAGTCAATATGGAGAGCTGCCACAAATAATCAGGCTTACCGCTCATTTCTCAGCAAAAT 2656
QY 257 -----V 257
DB 2657 CGACAGTCTTCGACCTTACCTAAGGATTCGATGATACCACTGATGTTGCTCTATA 2716
QY 257 aIleSerMetGlnGly-----TyrLysAlaAspAspTyrAla-----MetThrIleG 272
   ||||| ||| :|||:||||| ||||| |||
DB 2717 CCGACATATTAAGGAACTTACCTTACGCGAGATGATTTCTGCTCATTAACCGTTG 2776
QY 272 LysPheProGlySerThrAspArgTyrLeuThrSerTrpGlyValGlnAspArgIleGluA 292
   ||||| ||| :|||:||||| ||||| |||
DB 2777 AAGTTTCAGGT-----TATTGGCAGTTTGGGTGGCAGTGGGTGCTCCTCAGAA 2824
QY 292 sGlnLysAsnProArgIleGluValArgGlyIleLysGlnIle-----IleTrpLysG 310
   ||||| ||| :|||:||||| ||||| |||
DB 2825 ACCAA---GATCCCGGACCAAGGCTTCGAGACCAAGAGGTTGACGAGCTTTGTAAT 2881
QY 310 LuAlaMetSerAlaAspGlnAlaThr-----ArgI 320
   ||||| ||| :|||:||||| ||||| |||
DB 2882 CATCACAGCTCTTGTATTCACAGTCATTTACGAAGCTTCTCAAACTTCCAAAGATTGG 2941
QY 320 LeLysTyrAlaSerLysTyr-----AlaGlnSerAlaAsnTyrTrpLysA 335
   ||||| ||| :|||:||||| ||||| |||
DB 2942 TTAAGACACCAAGTCAGTACCAACCGTGTATTGCTCAAAATGCAAAACCTTCAAA- 3000
QY 335 snSerIleGlyMetLsnArgGlyLeuAlaArgLeuAspValIleGlyArg---LysArgA 354
   ||||| ||| :|||:||||| ||||| |||
DB 3001 -----GAGTGGGGAATCACCTTCTTGAAATTTGCGCCACGATGTGTCTA 3046
QY 354 LagIluGluArgLalPheAlaAspTrpIleArgLysAsnGly----- 367
   ||||| ||| :|||:||||| ||||| |||
DB 3047 GCCAAGACGCGCACTTCTTGATTTCTATCATTTGAATATGCTACGCGCTTGGAGACCGCT 3106
QY 368 -----LysSerAlaValTyrGly-----AspValLeuSerSerL 379
   ||||| ||| :|||:||||| ||||| |||
DB 3107 ACGATATTCGATGAGCAAGCAACAAATACGATCACTCAAAACCTCATGATGATGCC 3166
QY 379 euGlnLysAlaTyrLysGlnGlyAlaLysAlaAsnArgGluMetThrTyrLysSerGluT 399
   ||||| ||| :|||:||||| ||||| |||
DB 3167 TTCGTGCCCTTCAGCAGAGGATATCTCAGCATTCGCTGAC-----TGGGTTCCAGAAC 3220
QY 399 hLeuPheGly-----GlyThrGluValValArgPheAlaGlnPheAlaAsnAlaLeuA 417
   ||||| ||| :|||:||||| ||||| |||
DB 3221 AAATTTACATCTCCCTGGAAAAAGAGTGTACAA---GCTTCTGTACCAATACCTAGC 3277
QY 417 LaThrAsnProAspAlaHisAlaGlyIleLeuLysSerLeu---AspAspLysTyrLysA 436
   ||||| ||| :|||:||||| ||||| |||
DB 3278 GTACA---CCAGCTCCAAATATGCTGAATATCTACAAATAGCTTCACGACGCTAAACACGTA 3334

```

```

Oy 436 spTyLeuProserLeuAspArgLyValLeuProAlaMetLeuAspLysPheValArgGAG 456
    ::::: ||| ||| ||| ||| :::::
Db 3335 CCTTGGAAATGACTTCCAAAGGCAAGTACGGGAGTGCCCTTCTTATGAATGAAGGCA 3394
Oy 456 rgLleProAlaAspLysLeuProAspLlePheLysAsnValLleAspLysLysPheLysG 476
    :: ||||| ||||| ||| |||
Db 3395 AATACCCAGCA-----ATCTTGAGCGCGTGCA-GAT----- 3425
Oy 476 LyAspThrLysLTyrAlaAspPheValPheAspLysSerValLleAlaAlaArgAlaLleGlnAlaAs 496
Db 3426 -----TTCAAGAGGCGCGTAATTTACACCAATGAG 3456
Oy 496 ySPheAlaMetLeuLys--SerMetAspLysGluLysPheAlaLysAlaLleGluLys 515
    :: ||||| ||| ||| ||| ||| ||| |||
Db 3457 AAAATACGCAATGCTACGACCAAGTATTCATGTGACAGCAATATCCAAAGTACTGAG-- 3514
Oy 515 sAspProAlaValAlaLysSerLysSerValLleAlaAlaAlaArgAlaLleGlnAlaAs 535
Db 3515 -----CTCGCTAT-GTCTACACAGATTA 3536
Oy 535 PALMeAlaAsnAlaLysrAlaLleGluLysGlyLysArgLeuPhePheAlaLleArgLeu 555
    ::||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
Db 3537 CGCTACCAACCAATACTTACGCTCAAGCAGGTCAAACTTCCTTCTTAACAAATGAC 3596
Oy 555 gLMeLysrProGly-----ArgAlaLeuProSerAspAlaAsnPheThrMetArgMe 573
Db 3557 TGAATTACTGGAAGTGCGTTCCGTAGGGGTGAGATGAT-----GT 3638
Oy 573 tSerTyrGlySerLleLysGlyTyrGluPro-----GlnAspGlyAla-- 587
    :: ||||| ||| ||| ||| ||| ||| |||
Db 3639 CCAATACCTCTCAATGTGTGCTACCTCTCAAGATACCTTATTCACAACTGGGCGCAA 3698
Oy 588 -----TrpTyrAsnGlyrHisrThrLysGlyLysGlyValLeuGluLysGlnAspProLysSe 606
Db 3699 CCAGTGGATATCTTGATATAGAAATGGCAACATGTGCACAGGTGACGAGTGCATTTGATG 3758
Oy 606 rAspGluPheAlaValAlaGlnGluAsnLleLeuAspLysPheArgThrLysAsnTyrGlyAr 626
Db 3759 CAAGAAATCTTCTCTCTACAGCAATGGTCTCCAGCTA-----CGTACATGCTCTTG 3809
Oy 626 gTyrAlaGluAsnGlyLysLleuHisLle-----AlaPheLeuSe 639
    ::||| ::||| ::||| ::|||
Db 3810 CCAAGTAGTACGTCTCATGTGTATTATACATCCTTAAGGGGTTCAGGCCCTTTAACGG 3869
Oy 639 rAsnAsnAspLleThrGlyGlyLysAsnSerGlySerProValPheAspLysAsnGlyArgLe 659
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3870 ATTTATGATATTTGCGGGGCTCTCGCCACAGACGTTCCCTTCTTGATGGCAAGGTCATAT 3929
Oy 659 u 659
Db 3930 G 3930

RESULT 3
US-09-103-840A-1/c
: Sequence 1, Application US/09103840A
: Patent No. 6294328
: GENERAL INFORMATION:
: APPLICANT: FLEISCHMAN, Robert D.
: APPLICANT: WHITE, Owen R.
: APPLICANT: FRASER, Claire M.
: TITLE OF INVENTION: JOHN C.
: TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
: TITLE OF INVENTION: TUBERCULOSIS
: FILE REFERENCE: 24366-20007.00
: CURRENT APPLICATION NUMBER: US/09/103,840A
: NUMBER OF SEQ ID NOS: 2
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 1
: LENGTH: 4411529
: TYPE: DNA
: ORGANISM: Mycobacterium tuberculosis

```

OTHER INFORMATION: H37RV			
US-09-103-840A-1			
Alignment Scores:			
Pred. No.:	3.57e+03	Length:	4411529
Score:	111.00	Matches:	137
Percent Similarity:	29.45%	Conservative:	60
Best Local Similarity:	20.48%	Mismatches:	187
Query Match:	2.988	Indels:	286
Db:	4	Gaps:	35
US-10-008-355-2 (1-712) x US-09-103-840A-1 (1-4411529)			
QY	75	GIYLIETRRVAlSerASpBGLnglyLeuIlePheThraSnHtSHScysGIYTrGlyAla	94
DB	4023701	GGGCCACGGG-----CAAGGCTA-----CCGCAATACGGGCAT-----	4023666
QY	95	ILeGInSerGInSerThrValASpHtASpTYrLeuARqASpGlyPheValSerArgThr	114
DB	4023665	-----GTCGCGACAGGGGGGCGCCTTCGACATCTCGCGCGAGATGGTTCGCGGCGT---	4023615
QY	115	MeTgLyGInGluLeuProIleProGlyLeuSerValIstYrLeuARqLySLeVallys	134
DB	4023614	-----GGCTACAGACGTCCGCTTATCCCAACGTGACCGAC	4023579
QY	135	ValThraSpLySValGlyGlyGInLeuIleuYsglyLeThraSpGluMeTgLIARLeuArg	154
DB	4023578	ATCGCAACACAGATC-----	4023564
QY	155	LYsAlaGInGluValCySGLnGInLeuAlaLYsLYsGluASnAlaASpGluASnGInLeu	174
DB	4023563	-----CTGCGCAAGGCCCGCGCGGGC-----	4023540
QY	175	CysLIeValGInProPheTYr-----SerASnASnGluTYrPheLeuIleVal	190
DB	4023539	-----CGGCGGTGTGGAGTGGCGGCGCATCCAGACGCTGCTTCACCGCGCC	4023489
QY	191	TYrASpValPheLYsASpValARMeTValPheAlaProPProSerSerValGlyLySPh	210
DB	4023488	TACACGCTCTG-----GACGTC-----TTGCGCGCTGCCGCGAG-----	4023453
QY	211	GIYLIYASpThraSpASnTrPMeTTrPProARghIstHrGlyASpPheSerValPhe---	229
DB	4023452	-----CCGCGCGCCACCGGCAATACCATCCAGATGATC	4023420
QY	230	-----ArgValTYrAlaGlyAlaASpASnARgPro	239
DB	4023419	GAGATGATTCAGCGCCCTGATCCAGCGCGCACCGCGATACCGGTGGCTGAC-----	4023366
QY	240	AlaGluTYrSerLYsASpASnLYsProTYrLYsProValTYrPheAlaValaSer---	258
DB	4023365	-----GTTTACTTCGACGTTCTTGAGCTAC	4023342
QY	259	-----MeTGLnGlyTYrLYsAlaASpASpTYrAlaMeTThrILeGlyPhe	273
DB	4023341	CCGGAGTACGGCAGTATCTCGGTCACAAAGATCATCATCATCAGCGCGAAGGTGTG	4023282
QY	274	---ProGlySerThraSp---ArgTYrLeuThraSerTrPGLYValGluA-ASpArgILEG	291
DB	4023281	GCTCCCGGCAAGCGGGACACAGCGGCACTTCTGTGGAAAGGCGAAAGCCGGGTG--	4023224
QY	291	uASnGluASnASnProARgILEGlyValARgLYrLeYsGlnGlyIleTrPLysGluAl	311
DB	4023223	-AACCGTCGTGGCGCGACGCGTGGGGCGGGGGGCTCGGGGCTGGCATTCGG-----	4023173
QY	311	AmETSerAlaASpGlnAlaThraArgILEYsTYrAlaSerLYsTYrAlaGInSerAlaAS	331
DB	4023172	-----AATGCTCG	4023165
QY	331	nTYrTrP-----LYsASnSerILEGlyMeIAsnARgLYLeuAlaAR	345
DB	4023164	GCAATGGCTGCGACATATCTCGGGCGGAAATTCATATTCACATTGC-GGTGGAAAGCATTTT	4023106

US-08-446-137B-1

Alignment Scores:

Pred. No.:	0.0634	Length:	3279
Score:	110.00	Matches:	118
Percent Similarity:	31.59%	Conservative:	63
Best Local Similarity:	20.59%	Mismatches:	206
Query Match:	2.96%	Indels:	186
DB:	4	Gaps:	29

US-10-008-355-2 (1-712) x US-08-446-137B-1 (1-3279)

```

QY 29 TYPLeuLeuAsnGluLeuAsnGlnGluAsnLeuAspArgMetArgGluLeuGlyPheThr 48
    |||||
DB 1861 TGGTTATTAAAGATCTCTAAAGACAAAGCAATCAAGATTAAAGACAGCAATCACT 1920
QY 49 LeuProLeuAspSerLeuThrPheAspPysProSerTlleAlaAsnAlaValIle 68
    |||||
DB 1921 -----TCGATTTATATC--TTCAAGCTTAATCAATCAAGCAAAACAGTT----- 1962
QY 69 PheGlyGlyCysThrGlyIleThrValSerAspGlnGlyLeuIlePheThrAsnHis 88
    |||||
DB 1963 -----GAAGCGTAGAGACATTAAGACGAATCTTA-----AAAGCA 2001
QY 89 HisCysGlyTyrGlyAlaIleGlnSerGlnSerThrValAspHisAspTyrLeuArgAsp 108
    |||
DB 2002 CACGCTGGA-----GAAGAAACACACAGAA-----TTAAAGAT 2034
QY 109 GlyPheValSerArgThrMetGlyGlnGluLeuProIleProGlyLeuSerValIlystyr 128
    |||||
DB 2035 GGTATATGCACATATGATAGCAAGACAGACGCTTAAGAACTTTGAAAAATGATGAT 2094
QY 129 LeuArgIlyAlaValIlyThrAspIlyValGluGlyGln-----LeuIys 144
    |||||
DB 2095 GTTAAACACGACATAGCAATAGTTCAAGTGCAGACGGAAGATAGTATGATTAAG 2154
QY 145 ----GlyIleThrAspGluMetGluArgIlyValGlnGlnValIlyGln----- 161
    |||||
DB 2155 ATTTGAAGTTGCAGACGAAAGAACAGGTGAAGACATCTCCAGAGTTCAAGAGGTTAC 2214
QY 162 -----GluLeuAlaIlyLysGluAsnAlaAspGluAsnGlnIleCys 175
    |||||
DB 2215 GCACCTTACGAAAGACAGACGACGCTAAAGAACATTAAAGACATTAAGTT----- 2271
QY 176 IleValGluProPheTyrSerAsnAsnGluTyrPheIle----- 189
    |||||
DB 2272 -----AACATCGATACGAGTAGTACGATCAAGCTGACAGCGAAGA 2310
QY 190 ValTyrAspValPheIys---AspValArgMetValPheAlaProPheSerValGly 208
    |||||
DB 2311 TACTACTATGATTAATAAATCGAAGATTAAGATGAACACACAGTGAAGAACCGAGC 2370
QY 209 LysPheGlyGlyAspThr---AspAsnTyrMetTyrProArgIleThrGlyAspPheSer 227
    |||||
DB 2371 GAAACCCAGGAATCACAATTGATGATGTTTAAAGATGCTTAAGACAGAC----- 2424
QY 228 ValPheArgValTyrIleAlaIlyAspAsnArgProAlaGluTyrSerIysAspAsnIys 247
    |||||
DB 2425 -----GCATTCAAAGAAATTAAAGACGAGATCATCGTCTGAC----- 2463
QY 248 ProTyrLysProValTyrPheAlaIlyValSerMetGlnGlyTyrIlyAlaAspAspTyr 267
    |||||
DB 2464 -----ATATACCTTGTGATCAATC----- 2481
QY 268 AlaMetThrIleGlyPheProGlySerThrAspArgTyrLeuThrSerTyrPheValGlu 287
    |||||
DB 2482 -----AACAAAGCAAAACAGTAGAACGCTAGAA 2511
QY 288 AspArgIleGluAsnGlu-----AsnAsnProArg 297
    |||||
DB 2512 ----GGCTTAAAGACGAATCTTAAAGACACAGCTGAAAAAACCGAGCAAAACCCAGGA 2568
QY 298 IleGluVal-----ArgGlyIleLysGlnGlyIleThrPysGlu----- 310

```

```

DB 2569 ATTCACATTTGATGAATGTTATTAAAGATGCTAAAGACGTGCAATCAAGAAATTAA 2628
    |||
QY 311 -----AlaMetSerAlaAspGlnAlaThrArgIleIlyTyrAlaSerIysTyrAlaGln 328
    |||||
DB 2629 GAAGCAGAGATCATCTGCTCAATATCTATTCAACTTAATCAACAAAGCAAAACAGTAGAA 2688
QY 329 SerAlaAsnTyrTyrPysAsnSerIle-----GlyMetAsn 340
    |||||
DB 2689 GCGCTGATCATTTAAAGAACGAATCTTAAAGACACAGCTGAAAAACCGAGCGAAAC 2748
QY 341 ArgGlyLeuAlaIlyArgLeuAspValIleGlyArgIlyAspArgIleGluIlyArgAlaPheAla 360
    |||||
DB 2749 CCAGGATC-----ACAAATTGAT 2766
QY 361 AspTyrIleArgIlyAsnGlnIlyssSerAlaValTyrGlyAspValLeuSerSerLeuGlu 380
    |||||
DB 2767 GAATGGTTATTAAAGAACCTTAA-----GAA 2793
QY 381 LysAlaTyrLysGlnGlyAlaIlyAlaIlyAsnArgIlyMetThrTyrLeuSerGluThrLeu 400
    |||||
DB 2794 GATGCATTTAAAGAA-----TTAAAGAACGAGGAATTACTTGTACATATAC 2841
QY 401 PheGlyGlyThrGluValIlyArgPheAlaGlnPheAlaAsnAlaLeu----- 416
    |||||
DB 2842 TTTGATGCTATCAACCAAAACCTATTGAAGCGCTGAGGATTAAGAAATGAAATC 2901
QY 417 -----AlaThrAsnProAspAlaHisAlaGlyIleLeuIlySerLeuAspIlystyr 434
    |||||
DB 2902 TTTAAAGCTCATATAAAGATGAGAACCAAGTAAACCAAGGAGACAAACCA 2961
QY 435 LysAspTyrLeuProSerLeuAspArgIlyValLeuProAlaMetLeuAspIleValArg 454
    |||||
DB 2962 GAAATTAATAAACCAGCTGAAGATTAATAA----- 2991
QY 455 ArgArgIleProAlaAspIlyLeuProAspIlePheIlyAsnValIleAspIlystyrPhe 474
    |||||
DB 2992 -----CCAGAACACAAAAACCTGTTGAAGATTAATAAACCAGAACCAAAAAA--- 3039
QY 475 LysGlyAspThrLysIlyTyrIleAspPheValPheAspIlySerValIleProTyrSer 494
    |||||
DB 3040 CCAGCTAAACAGATTA----- 3057
QY 495 AspIlyPheHisAlaMetLeuLysSerMetAspIlyGluIlyPheAlaIlyAlaIleGlu 514
    |||||
DB 3058 -----GATTCCTCCAAATTAAGAAAAAGCTTAATTA----- 3090
QY 515 LysAspProAlaValGluLeuSerIlyssSerValIleAlaIlyAlaIlyArgAlaIleGlnAla 534
    |||||
DB 3091 -----CCAAAAGCTGTGACGAGAGCTGAATCTTAACATTGACAGACGACCTTTATCA 3144
QY 535 AspAlaMetAlaAsnAlaTyrAlaIleGluIlyGlyIys 547
    |||||
DB 3145 ACAGCAGCAGGTGCTTACGTTTCACCTTAATAAACCTTAA 3183

```

RESULT 5

```

US-09-103-840A-2/C
: Sequence 2, Application US/09103840A
: Patent No. 6294328
: GENERAL INFORMATION:
: APPLICANT: FLEISCHMAN, Robert D.
: APPLICANT: WHITE, Owen R.
: APPLICANT: FRASER, Claire M.
: APPLICANT: VENTER, John C.
: TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
: FILE REFERENCE: 24366-2007.00
: CURRENT APPLICATION NUMBER: US/09/103.840A
: CURRENT FILING DATE: 1998-06-24
: NUMBER OF SEQ ID NOS: 2
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 2
: LENGTH: 4403765

```

```

; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Alignment Scores:
Pred. No.:      5.91e+03      Length:      4403765
Score:          109.00        Matches:      136
Percent Similarity: 29.34%      Conservative: 60
Best Local Similarity: 20.36%      Mismatches:  189
Query Match:      2.93%        Indels:      284
DB:               4          Gaps:      34

US-10-008-355-2 (1-712) x US-09-103-840A-2 (1-4403765)
QY      75  gLyIleThrValSerAspSngIngluLeuIlePheThrAsnHisScySglYrGlyAla 94
      ||| ||||| ||||| ||||| ||| |||||
Db 4015782 GGTGGCACGCTG-----CAGGCGCTA-----CCGCATATCGGGCAT----- 4015747

QY      95  IleGlnSerGlnSerThrValAspHisAspTyrLeuArgAspGlyPheValSerArgThr 114
      ::||| ||| ||| ||||| ||||| |||||
Db 4015746 -----GTCCCGCAGCGGGCTGCTTCGACATCTGCCCGGATGGTGGCTCGCGGT---- 4015696

QY      115 MetGlyGlnGluLeuProIleProGlyLeuSerValLysTyrLeuArgGlySllValLys 134
      ||| ||| ||| ||||| ||||| |||||
Db 4015695 -----GGCTACGACGTCGCGCTTATCCGCAACGTGACCGCAG 4015660

QY      135 ValThrAspLysValGlnGlyGlnLeuLysGlyIleThrAspGlnMetGluArgLeuArg 154
      ::||| ||||| ||||| ||||| |||||
Db 4015659 ATCGAAGACAAATC----- 4015645

QY      155 LysAlaGlnGluValCysGlnGluLeuAlaLysGlnAsnAlaAspGlnAsnGlnLeu 174
      ||||| ||||| ||||| |||||
Db 4015644 -----CTGGCCAGCGCGCGCGCGGCGC----- 4015621

QY      175 CysIleValGluProPheTyr-----SerAsnAsnGlnLysPheLeuIleVal 190
      ||||| ||||| ||||| |||||
Db 4015620 -----CGCGCGGGGGAGTGGCGGCTACCCAGCAGCGTCTCACCGCGGCC 4015570

QY      191 TyrAspValPheLysAspValArgMetValPheAlaProProSerSerValGlyLysPhe 210
      ||||| ||||| ||||| ||||| |||||
Db 4015569 TACGACGCTCTG---GAGCTC-----TTGCCCGCTCCGCGAG----- 4015534

QY      211 GlyGlyAspThrAspAsnTrpMetTyrProArgHisThrGlyAspPheSerValPhe--- 229
      ||||| ||||| ||||| |||||
Db 4015533 -----CCGCGCGCCACCGCGCATATCACCCAGATGATC 4015501

QY      230 -----ArgValTyrAlaGlyAlaAspAsnArgPro 239
      ||| ||| ||| |||
Db 4015500 GAGATGATCGACGCCCTGATCCAGCGCGGCCAGCGGTATACCGGTGGCGGTAC----- 4015447

QY      240 AlaGluTyrSerLysAspAsnLysProTyrLysProValTyrPheAlaAlaValSer--- 258
      ||||| ||||| ||||| |||||
Db 4015446 -----GTTACTTGTGACCTTGTGACCTAC 4015423

QY      259 -----MetGlnGlyTyrLysAlaAspAspTyrAlaMetThrIleGlyPhe 273
      ::||| ||||| ||||| |||||
Db 4015422 CCGAGATACGGCCAGTTCCTGACAAAGATCGATGATGCATCAGGCGCAAGGTGTG 4015363

QY      274 ---ProGlySerThrAsp---ArgTyrLeuThrSerTrpGlyValGlu---AspArgIleG 291
      ||| ||| ||| ||| ||| ||| |||
Db 4015362 GCTGCCGCGCAACGCGGACACGCGACCTTCACCTTTGTGAAAGCGCAAAAGCGGGGTG-- 4015305

QY      291 uAsnGlnAsnAsnProArgIleGluValArgGlyIleLeuSngIngluIleTrpLysGluAl 311
      ||||| ||||| ||||| ||||| |||||
Db 4015304 -AACGCTGCTGGCGACGCGGTGGGCGCGCGGCGTCCGGGTGGATCTGG----- 4015254

QY      311 aMetSerAlaAspGlnAlaThrArgIleLysTyrAlaSerLysTyrAlaGlnSerAlaAs 331
      ::||| ||||| ||||| |||||
Db 4015253 -----AATGCTCG 4015246

```

```

QY      331 nTyrTrp-----LysAsnSerIleGlyMetAsnArgGlyLeuAlaIar 345
      ||| ||||| ||||| ||||| |||||
Db 4015245 GCATATGCTCCGACGACTATCTCGCGCGGAGATTCGATTCATCTGC- GGTGATGATTT 4015187

QY      345 gLeu-----AspValIleGlyArgLysArgAlaGluGluArgAlaPheAl 360
      ::||| ||||| ||||| ||||| |||||
Db 4015186 AGCTTCTCCCGCATCATGAGAAAGAAATGCCAGAGCCGCGCGCGGTATGGGTTCG 4015127

QY      360 a-----AspTrpIleArgLysAsnGlyLys-----SerAlaValTyr 372
      ||||| ||||| ||||| ||||| |||||
Db 4015126 CCGTACTGTGCTGCACAAATGCTGGGTGACGATGGCGGGAGAAAGATGCAATGCGCT 4015067

QY      372 rGlyAspValLeuSer-----SerLeuGlnLysAlaTyr 383
      ||||| ||||| ||||| ||||| |||||
Db 4015066 GGGCAACGTTGTTCACATCCGCGGATGTTGCAGCGGGTGGCGCGCTGACCTCGCTTA 4015007

QY      383 rLysGlnGlyAlaLysAlaAsnArgGlnMetThrTyrLeuSerGlnLysPheGlyG 403
      ||||| ||||| ||||| ||||| |||||
Db 4015006 TTACTGGGCGACGCGGACACTACCGGTGATGCTCGAGTTCTCGAGAC----- 4014958

QY      403 YThrGluValValArgPheAlaGlnPheAlaAsnAlaLeuAlaThrAsnProAspAlaH 423
      ||||| ||||| ||||| ||||| |||||
Db 4014958 ----- 4014958

QY      423 sAlaGlyIleLeuLysSerLeuAspLysTyrLysAspTyrLeuProSerLeuAspAr 443
      ::||| ||||| ||||| ||||| |||||
Db 4014957 -----GCTATCAGAGATCGGCTGAAGCTTATGTGCGG----- 4014925

QY      443 gLysValLeuProAlaMetLeuAspIleValArgArgArgIle-----ProAl 459
      ||||| ||||| ||||| ||||| |||||
Db 4014924 -----CTAAGAGACTTCTCTACCCGCTTGACACCGCGGTGGCGCGCTGCGCCGG 4014872

QY      459 aAspLysLeuProAspIlePheLysAsnValIleAspLysLysPheLysGlyAspThrLy 479
      ||||| ||||| ||||| ||||| |||||
Db 4014871 CGATCCAGCCCA----- 4014859

QY      479 sLysTyrAlaAspPheValPheAspLysSerValProTyrSerAspLysPheHisAl 499
      ||||| ||||| ||||| ||||| |||||
Db 4014859 ----- 4014859

QY      499 aMetLeuLysSerMetAspLysGlnLysPheAlaLysAlaIleGluLysAspProAlaVa 519
      ::||| ||||| ||||| ||||| |||||
Db 4014858 -----CGTTCGCCGAAGCGCTCGACGACGACTGTCTGCT 4014824

QY      519 lGluLeuSerLysSerValIleAlaAlaArgAlaIleGlnAlaAspAlaMetAlaAs 539
      ::||| ||||| ||||| ||||| |||||
Db 4014823 TCCGATCCGCGCTCGCGGAGATTACACACGTCGGCGGCG----- 4014787

QY      539 nAlaTyrAlaIleGluLysGlyLysArgLeuPhePheAlaGlyLeuArgGlnMetTyrPr 559
      ::||| ||||| ||||| ||||| |||||
Db 4014786 -----GAAGGCAACCGGGGCTCGATGCGCGGCGACACGAC----- 4014751

QY      559 ogLysArgAlaLeuProSerAspAlaAsnPheThrMetArgMetSerTyrGlySerIleLy 579
      ||||| ||||| ||||| ||||| |||||
Db 4014750 -----GGGGCTTGCGAAGCGCTAGGGCGATGTGGGGCGATGATG----- GGCATCTCT 4014704

QY      579 sGlyTyrGluProGlnAspGlyAlaTrpTyrAsnTyrHisThrThrGlyLysGlyValLe 599
      ::||| ||||| ||||| ||||| |||||
Db 4014703 GGGCTGTGACCCGCTCGACACGCTCG----- 4014676

QY      599 uGluLysGlnAspProLysSerAspGluPheAlaValGlnGluAsnIleLeuAspLeu-- 618
      ::||| ||||| ||||| ||||| |||||
Db 4014675 -GAATCCCGAGCGAAACTCTCGGACAGCGTGGCCCGCGCTGCGATGTGCTGCTCAGAGTGA 4014617

QY      619 -PheArgThrLysAsnTyrLeuArgTyrAlaGluAsnGlnLeuHisIleAlaPheLe 638
      ::||| ||||| ||||| ||||| |||||
Db 4014616 ACTACAGAAATCGGAAGAGCCCGCGAGCGACGCACTGTGGGGCTC----- 4014571

QY      638 uSerAsnAsnAspIleThrGly-----GlyAsnSerGlySerProValPheAspLysAs 656
      ::||| ||||| ||||| ||||| |||||
Db 4014570 -----GCCAGAGATTCGCGGTGCGCTGAAAGAGACCGGCATCGAGTCCACCGAC----- 4014520

```


OY 656 nglYargleuilegIleuAlaPheAspGlyAsnTrpGluAlaMetSerGlyAspIleGI 676
DB 4014519 -----ACCGCCGAGGGGCCACAGTGGCTGCTGGGTGTGACACCAA 4014476
OY 676 uPheGluProAspLeuGlnArg 683
DB 4014475 GTAGATGCCCGTAACCTCTCGG 4014454

RESULT 6
US-08-335-844A-7
; Sequence 7, Application US/08335844A
; Patent No. 6066503
; GENERAL INFORMATION:
; APPLICANT: GRAHAM, MARGARET
; APPLICANT: SMITH, TREVOR STANLEY
; APPLICANT: MURN, EDWARD ALBERT
; APPLICANT: KNOX, DAVID PATRICK
; APPLICANT: OLIVER, JOANNA JANE
; APPLICANT: NEWTON, SUSAN ELIZABETH
; TITLE OF INVENTION: RECOMBINANT DNA MOLECULES ENCODING
; TITLE OF INVENTION: AMINOPEPTIDASE ENZYMES AND THEIR USE IN THE PREPARATION OF
; TITLE OF INVENTION: VACCINES AGAINST HELMINTH INFECTIONS
; NUMBER OF SEQUENCES: 73
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rothwell, Figg, Ernst & Kurz
; STREET: Suite 701-E, 555 Thirteenth St., N.W
; CITY: Washington
; STATE: D. C.
; COUNTRY: U.S.A.
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/335, 844A
; CLASSIFICATION: 424
; FILING DATE: 09-JAN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB PCT/GB93/00943
; FILING DATE: 06-MAY-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 920936
; FILING DATE: 08-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: WALKER, Barbara W.
; REGISTRATION NUMBER: 35,400
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)783-6040
; TELEFAX: (202)783-6031
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2472 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-335-844A-7

Alignment Scores:
Pred. No.: 0.113 Length: 2472
Score: 106.00 Matches: 111
Percent Similarity: 33.15% Conservative: 71
Best Local Similarity: 20.22% Mismatches: 230
Query Match: 2.85% Indels: 138
DB: 3 Gaps: 22
US-10-008-355-2 (1-712) x US-08-335-844A-7 (1-2472)

OY 89 HScYsGlyTYrGlyAlaIleGIInSerGlnSerThrValAspHisAspTYrLeuArgAsp 108
DB 283 CATCATGCTTCGCAATCTGGTACACTGAAGTGGTGGATGATACGTGTGAACGAA 342
OY 109 GlyPheValSerArgThrMetClyGluGluLeuProIleProGlyLeuSerValLysTYr 128
DB 343 GGTTTTGA-----ACATTGTTGATGATCTTGGAAAGACGAAATTAGCCACAACAT 396
OY 129 LeuArgLysIleValLysValThrAspLysValGluGlyInLeuLysGlyIleThrAsp 148
DB 397 TTCACA-----ACCGAAGATTCTCTTGGCTGCATGCAATGCAATCGCGGAATGAGACT 450
OY 149 GluMetGlu-----ArgLeuArgLysAlaGlnGluValCys 160
DB 451 GACGCGGAGCATCGAGCATCCGCTTGTGATTGAGATTGACAAAGCCGAGAAAGTTGCC 510
OY 161 Gln-----Glu 162
DB 511 GAAGCCTTTGACGATATTTCATACGCCAAGGAGCGTCAGTTCTCATAGCTACGGGCT 570
OY 163 LeuAlaLysLysGluAsnAlaAspGluAsnGlnLeuCysIleValGluProPheTYrSer 182
DB 571 TTGATTGGAGAGACAAATTACAGAAATGCTGTTGCAATACCTCAAGAGTTCTCTAC 630
OY 183 AsnAsnGluTYrPheLeuIleValTYrAspValPheLysAsp---ValArgMetValPhe 201
DB 631 AGCAATGCACAAGCAGCGATCTGTGAAACGCTTCAATGAATGTAAGAGGTGTAA 690
OY 202 AlaProProSerSerValGlyLysPheGlyAspThrAspAsnTrpMetTrpArg 221
DB 691 GGTCTGACGGCAACGTCATGAAATTCAGCAATTACCAGACAGTCAATGATGATG 750
OY 222 HStrHglYasPheSerValPheArgValTYrAlaGlyAlaAspAsnArgProAlaGlu 241
DB 751 GGTATCTCTGTCGTTAACTAGAACAAATTATATGACCCGCTTAAGCTTACGAGAGC 810
OY 242 TYrSerLysAspAsnLysProTYrLysProValTYrPheAlaAlaValSerMetGlnGly 261
DB 811 CGGTACAGACAAATTAAGAC-----GCCTTGAAACACAGAA 849
OY 262 TYrLysAlaAspAspTYrAlaMetThrIleGlyPhePro-----GlySer 276
DB 850 TATCGTAATCCAAATATACGGGTTCAAGTGGATGTTCCCTATGATACAGAAACGAT 909
OY 277 ThrAspArgTYrLeuThrSerTrpGlyValGluAspArg-----IleGluAsn 292
DB 910 ACCAAGAGGTGAACGCAACATGCTTAAGAAAGATGAACCCCTACTTGAACTCAAC 969
OY 293 GluAsnAsnProArgIleGluValArgGlyIleLysGlnGlyIleTrpLysGluAlaMet 312
DB 970 AATCGGATACATCCCTGTGTGTAACCTGATCGACATGATTTATTCGACAA----- 1023
OY 313 SerAlaAspGlnAlaThrArgIleLysTYrAlaSerLysTYrAlaGlnSerAlaAsnTYr 332
DB 1024 -----AACTAATGATCCCAACGCT 1041
OY 333 TrpLysAsnSerIleGlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLys 352
DB 1042 TCGAAAGA-----ATATCAAGCAGCTC 1065
OY 353 ArgAlaGluGluArgAlaPheAlaAspTrpIleArgLysAsnGlyLysSerAlaValTYr 372
DB 1066 AAGAAAGATCAACAGCTTCGTCCGTCACAGCAAG-----AACGCTATCATTA 1113
OY 373 GlyAspValLeuSerSer-----LeuGluLysAlaTYrLys--- 384
DB 1114 AGCGATGATTTGTCGACAGTACAGATTGACCGAATGCACTATGAAGATTCATCAACTA 1173
OY 385 ---GluGlyAlaLysAlaAsnArgGluMetThrTYrLysSerGluThrPheGlyGly 403
DB 1174 CTTCATATATCCAAATAAGAGAGAAATTTGCTTGGAGGAAGAGCTGTGCGGCAATG 1233
OY 404 ThrGluValValArgPheAlaGlnPheAlaAsnAlaLeuAlaThrAsnProAspAlaHis 423


```

Db 1234 TTGGCAGTTTAAAGTTC-----TTGGTAATAGCCGAGACAAACCA---GCTAGA 1284
QY 424 AlAclYlleuLysSerLeuAspAspLysTyrLysAspTyrLeuProSerLeuAspArg 443
Db 1285 GCTTACATGATGACCATATTGAACCGATGATATATAAG-----ACAGCATGTGATTAC 1338
QY 444 LysValleuProAlaMetLeuAspLleValArgArgArgLleProAlaAspLysLeuPro 463
Db 1339 ATCTGTCAAGATTATTGTGATGATACGTTTATTCACAAAATAATTAATCTCAAAA- 1392
QY 464 AspLlePheLysAsnValLleAspLysLysPheLysGlyAspThrLysLysTyrAlaAsp 483
Db 1393 GATATCATTTGATGATTTGTTCCTTGATCAAAAGACGTATATAAGCATATATAAGAT 1452
QY 484 PheValPheAspLysSerValValProTyrSerAspLysPheLysAlaMetLeuLysSer 503
Db 1453 ATCTTCTACGATGAG-----GTTATGCTCAAGTGT 1482
QY 504 MetAspLysGlyLysPheAlaLysAlaLleGlyLysAspProAlaValGluLeuSerLys 523
Db 1483 AAGCCCGGAGACGACGACCAAAATGCGTTAAG----- 1515
QY 524 SerValLleAlaAlaAlaArgAlaLleGlyAlaAspAlaMetAlaAsnAlaTyr----- 541
Db 1516 -----GTTCCGCTCTCTCTGAGCCCATGTTACTGTTAT 1551
QY 542 -----AlaLleGlyLysGlyLysArgLeuPhePheAla-- 552
Db 1552 GGTGTACAGAGAGGTGTGAAGAAGCTTTGAAAAGTGATGAGGCGTATCTAGCAGAA 1611
QY 553 -----GlyLeuArgGlyMetTyrProGlyArgAlaLeu-----ProSerAspAlaAsn 568
Db 1612 GATTTCAACTGGA-GAAGGATATCTGTTCAAAAGCTTGCGATGCCAACAAGATGTTAC 1670
QY 569 PheThrMetArgMetSerTyrGlySer 577
Db 1671 AGCTTAAGAAGACTTCTTTTGCAGC 1697

RESULT 7
US-08-335-844A-20
; Sequence 20, Application US/08335844A
; Patent No. 6066503
; GENERAL INFORMATION:
; APPLICANT: GRAHAM, MARGARET
; APPLICANT: SMITH, TREVOR STANLEY
; APPLICANT: MUNN, EDWARD ALBERT
; APPLICANT: KNOX, DAVID PATRICK
; APPLICANT: OLIVER, JOANNA JANE
; APPLICANT: NEWTON, SUSAN ELIZABETH
; TITLE OF INVENTION: RECOMBINANT DNA MOLECULES ENCODING
; TITLE OF INVENTION: AMINOPEPTIDASE ENZYMES AND THEIR USE IN THE PREPARATION OF
; NUMBER OF SEQUENCES: 73
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rothwell, Figs. Ernst & Kurz
; STREET: Suite 701-E, 555 Thirteenth St., N.W
; CITY: Washington
; STATE: D. C.
; COUNTRY: U.S.A.
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/335,844A
; FILING DATE: 09-JAN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB PCT/GB93/00943
; FILING DATE: 06-MAY-1993

```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9209936
; FILING DATE: 08-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: WALKER, Barbara W.
; REGISTRATION NUMBER: 35,400
; REFERENCE/DOCKET NUMBER: 1181-223A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)783-6040
; TELEFAX: (202)783-6031
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3358 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEetical: NO
; ANTI-SENSE: NO
; US-08-335-844A-20

Alignment Scores:
Pred. No.: 0.181 Length: 3358
Score: 106.00 Matches: 103
Percent Similarity: 33.66% Conservative: 71
Best Local Similarity: 19.92% Mismatches: 223
Query Match: 2.85% Indels: 120
Gaps: 21

US-10-008-355-2 (1-712) x US-08-335-844A-20 (1-3358)
QY 89 HisCysGlyTyrGlyAlaLleGlnSerGlnSerThrValAspHisAspTyrLeuArgAsp 108
Db 1169 CATGAGTGTTCGCGATTCGCTCACACGTGAAGTGTGGATGATACGTGGTGAACGAA 1228
QY 109 GlyPheValSerArgThrMetGlyGluGluLeuProLleProGlyLeuSerValLysTyr 128
Db 1229 GGTTTTGA-----ACATTGTGAGTATCTTGATGATGAGCAAAATTAAGCAACAAT 1282
QY 129 LeuArgLysLleValLysValThrAspLysValGluGlyGluLeuLysGlyLleThrAsp 148
Db 1283 TTCAGA-----ACGCAAAATTTCTTCTGTCGATGATGATGCCGGAATGAGACT 1336
QY 149 GluMetGlu-----ArgLeuArgLysAlaGlnGluValCys 160
Db 1337 GACTCGGACAGCATGAGCCATCCGTTTCTGTTAGATGACAAAGCGGAGAGTGGC 1396
QY 161 Gln----- 162
Db 1397 GAAACCTTTGACGATATTTCATACGCCAAGGAGCGTCAGTTCTCACTATCTACGGGCT 1456
QY 163 LeuAlaLysLysGluAsnAlaAspGluAsnGlnLeuCysLleValGluProPheTyrSer 182
Db 1457 TTGATTGAGAGACAAATTACAGGAATGCTGTGCAATACCTCAAGAAAGTTCTCTAC 1516
QY 183 AsnAsnGlyTyrPheLeuLleValTyrAspValPheLysAsp---ValArgMetValPhe 201
Db 1517 AGCAATGCACAGACGCGCATCTGTGAACGCTTCAATGAAGTTGCAAAAGTGTAAAG 1576
QY 202 AlaProPheSerValGlyLysPheGlyLysAspThrAspAsnTyrPmetTyrProArg 221
Db 1577 GGTCTGACGCGCAACGTCATGAAATATGACCAATTATACCATGCTGAGCTATACATG 1636
QY 222 HisThrGlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgProAlaGlu 241
Db 1637 GGTATCTCTGTGTTAAAGTGAAGAAATTAATCGACCGCCCTAAAGGTTACACAGC 1696
QY 242 TyrSerLysAspAsnLysProTyrLysProValTyrPheAlaValSerMetGlnGly 261
Db 1697 CGGTACACAGACAAATTAAGAC-----GCCTTGGAACACAGACAAA 1735
QY 262 TyrLysAlaAspAspTyrAlaMetThrLleGlyPhePro-----GlySer 276

```

```

Db 1736 TATGTAATCAAAATACGGGTTCAAGTGGAGTTCCTCCCTATGATACAGAGGCAAT 1795
QY 277 ThrAsparGlyLeuThrSerTrpGlyValGluAsparG-----IleGluAsn 292
Db 1796 AGCAAGAGGTGAACGACATGGCTAAAGATGACCCGCTACTTGAACGTCAC 1855
QY 293 GluAsnAsnProArgIleGluValArgGlyIleLysGlnGlyIleTrpLysGluIleMet 312
Db 1856 AATCGGATACATCCCTTGTGTGAACGCTGACATGATGATTTATGCACAA----- 1909
QY 313 SerAlaAspGlnAlaThrArgIleLysTrpAlaSerLysTrpAlaGlnSerAlaAsnTrp 332
Db 1910 -----AACATGATGCCAACGGT 1927
QY 333 TrpLysAsnSerIleGlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLys 352
Db 1928 TGCAGAAAG-----ATAATCAAGCAAGCTC 1951
QY 353 ArgAlaGluGluArgAlaPheAlaAspTrpIleArgLysAsnGlyLysSerAlaValTrp 372
Db 1952 AAGAAAGATCACAAGGCTTCGCTCCAGACAAAG-----AACGCTATCATTA 1999
QY 373 GlyAspValLeuSerSer-----LeuGluLysAlaTrpLys--- 384
Db 2000 AGCGATGCAATTTCTGCTGCTACGATGACCAATGCACTATGAACTGATTCGAACCTA 2059
QY 385 ---GluGlyValAlaLysAlaAsnArgGluMetThrTrpLeuSerGluThrLeuPheGlyGly 403
Db 2060 CTTCGAATATGCCAAAATAGACAGGAAATTTCTTCCTTGGACAGAGCTCTGTCGGCATG 2119
QY 404 ThrGluValValArgPheAlaGlnPheAlaAsnAlaLeuAlaThrAsnProAspAlaHis 423
Db 2120 TTCGAGTTTAAAGTTC-----TTCGTAATGAGCCGAGACAAACCA---GCTAGA 2170
QY 424 AlaGlyIleLeuLysSerLeuAspLysTrpLysAspTrpLeuProSerLeuAspArg 443
Db 2171 GCTTACATGATGAGCATATTATTAACCGATGATTAATTAAG-----ACGACATTTGATTAC 2224
QY 444 LysValLeuProAlaMetLeuAspIleValArgArgGlyLeuProAlaAspLysLeuPro 463
Db 2225 ATGCTCAAGAAATATTGGATGATACGTTATTCACAAAATTAATCTCAAAAG----- 2278
QY 464 AspIlePheLysAsnValIleAspLysLysPheLysGlyAspThrLysLysTrpAlaAsp 483
Db 2279 GATATCATGATCATATTGCTCCCTGATCAAGACAGCTGATAAGCAATATAAGAT 2338
QY 484 PheValPheAspLysSerValAlaProTrpSerAspLysPheHisAlaMetLeuLysSer 503
Db 2339 ATCTTCTACATAG--GTTATGCCAAGTGAAGCCGGGAGACGACCAACCAATGC 2395
QY 504 Met-----Asp 505
Db 2396 GTTAAGTTCCGCTCCTCTTCGAGCAATGTTTACTGTTATGCTGACAGAAAGTGT 2455
QY 506 LysGluLysPheAlaLysAla-----IleGluLysAspProAlaValGluLeu 521
Db 2456 GAAGAAGCTTTGAAAAGGATGGATGGCTGTATCTAGCAGAAAGAT-----GTTCAACTG 2509
QY 522 SerLysSerValIleAlaAlaAlaArgAlaIleGlnAlaAspAlaMetAla 538
Db 2510 GAGAAAGGATATCTGTTCAAGCCTTGGCATGCCACAAAGATGTTACAGCT 2560

```

```

; CURRENT FILING DATE: 1999-06-08
; PRIOR APPLICATION NUMBER: US 08/714,402
; PRIOR FILING DATE: 1996-09-16
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 1
; LENGTH: 3698
; TYPE: DNA
; ORGANISM: SFBP gene
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (196)..(3681)
; US-09-327-536-1

Alignment Scores:
pred. No.: 0.211 length: 3698
Score: 106.00 matches: 90
Percent Similarity: 35.12% conservative: 74
Best Local Similarity: 19.27% mismatches: 140
Query Match: 2.85% indels: 163
DB: 4 gaps: 24

US-10-008-355-2 (1-712) x US-09-327-536-1 (1-3698)
QY 165 LysLysGluAsnAlaAspGluAsnGlnLeuGlyIleValGluProPheTrpSerAsnAsn 184
Db 186 GAGAAATATTAATACACCAAAAATAGCTAATAGTTAAGCTTCTGTTATCCCTAACAG 245
QY 185 GluTrpPheLeuIleValIleLysAspValPheLysAspValArgMetValPheAlaPro 204
Db 246 ATTTATTTAGGTTTATTTATGCTTTTATAGCATGTCGCG----- 287
QY 205 SerSerValGlyLysPheGlyGlyAspTrpAsnTrpMetTrpProArgHisThr-Gl 224
Db 288 ---AGTATCAAGAGCATGCCGA-----AACAGAAATGCGCAACAAACAAAG 335
QY 224 LysPheSerValPheArgValTrpAlaGlyAlaAspAsnArgPro----- 239
Db 336 ATCTTTTGAATACAGAAAGTGC-----GACCAAAACAAATTAAGCCTTACCGGAGCAAC 389
QY 240 ---AlaGluTrpSerLysAspAsnLysProTrpLysProValTrpPheAlaValSe 258
Db 390 TTCTTCTACGACATCAAGATGGCAAG-----GGAACATC 425
QY 258 TrpMetGlnGlyTrpLysAlaAspAspTrpAlaMetThrIleGlyPheProGlySerThrAs 278
Db 426 TGTTCAAAGCTTCACCTCAATGATTAAGTAT----- 459
QY 278 ParGlyTrpLeuThrSerTrpGlyValGluAspArgIleGluAsnGluAsnProArgIle 298
Db 460 -----GT 461
QY 298 eGluValArgGlyIleLysGlnGlyIleTrp-----LysGluAlaMetSerAlaAsp 316
Db 462 AGATGCTCAAAATCTCCACACGAGGACTTATACCTTAAAGAAAGAAACAGACCAAT-- 519
QY 316 nAlaThrArgIleLysTrpAlaSerLysTrpAlaGlnSerAlaAsnTrpTrpLysAsnSe 336
Db 520 -----GTTATGATAAACCACGCGGACTTGACACTGC 554
QY 336 rIleGlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGlu 356
Db 555 TGTATTATGAGAAC--GGCTATACCAAGTTG-----GTTCAAAA 590
QY 356 ValArgAlaPheAlaAspTrpIleArgLysAsnGlyLysSerAlaValTrpGlyAspVal 376
Db 591 TCCCTATTAATGGGAAATCATCATTAACAGAGGTCACAA-----GATGTTAG 638
QY 376 uSerLeuGluLysAlaLysTrpLysGluGlyAlaLysAlaAsnArgGluMetThrTrp 396
Db 639 TACTTTTACAG-----TTGAAATATCCCAAAATGTCACTGT 677
QY 396 uSerGluThrLeuPheGlyGlyThrGluValValArgPheAlaGlnPheAlaAsnAlaLe 416

```

```

Db 678 TTTAAA-----TATGGAAACAGAGGTT-----AGTAGTGGCGC 713
Qy 416 uAlaThrAsnProAspAlaHisAlaGlyIleLeuLys---SerLeuAspSplysTyrLy 435
Db 714 AGCGGATTTCTACCGCAACATGCGCCCTATTAAATGCTCTTTAGAGTTGAACAAA 773
Qy 435 sasp-----TyrLeuProSerLeuAspArgLysVa 445
Db 774 GGATTAATCTGAACAATCAACCCAGGTGATACCTTTGTGTACAGCTGATGAGCTCT 833
Qy 445 lLeuProAlaMetLeuAspIleValArgArgAlaIleProAlaAspLysLeuProAsp 465
Db 834 CAATCCT----- 840
Qy 465 ePheLysAsnValIleAspLysLysPheLysGlyAspThrLysLysTyrAlaAspPheVa 485
Db 841 -----AAAGGTATCATGCAAGATATCCCTTAATATCAT 872
Qy 485 lPheAspLysSerValValProTyrSer---AspLysPheHisAla----- 499
Db 873 TTACGACAGTGCATATAGTCCGCTTGGGATTTGGAATATACCATGCTGAGAACATCACT 932
Qy 500 -----MetLeuLysSerMetAspLysGlyLysPheAlaLysAla11 513
Db 933 TATCTATACCTTTCACAGATATATTTGCGGGTTTATGATATA----- 972
Qy 513 eGluLysAspProAlaValAlaGluLeuSerLysSerVal-----lLeAlaAlaAla1r 530
Db 973 -----GTCCAGTTGCTCGCAAGATTGAGCTTATCTTCAAGATAAGAA 1016
Qy 530 gAlaAlaGlnAlaAspAlaMetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPh 550
Db 1017 AGTGTGGAAATATCTGATCTCAATTTTAAGAGTACCATAGTGGCGGAGGACATCAC 1076
Qy 550 ePheAlaGlyLeuArgGluMetTyrProGlyArgAlaLeuProSerAspAlaAsnPh-- 569
Db 1077 CTATTAAGGAACGGTATGATGTTCTTTATGGAATAGAGACACATTAAGAAAGCAATTTAT 1136
Qy 570 -ThMetArgMetSerTyr-----GlySerIleLysGlyTyrGluProGlnAspGly-- 586
Db 1137 TACTAATGGAATGAGCAATGTGGGTGGAGTATTTGAAGCTACAAACCAACGAGGAGA 1196
Qy 587 -----AlaTrpTyrAsnTyr 591
Db 1197 ATTTGTCTGTATGTTAT 1215

RESULT 9
US-08-600-982-22
; Sequence 22, Application US/08600982
; Patent No. 6120991
; GENERAL INFORMATION:
; APPLICANT: Carter, William G.
; APPLICANT: G11, Susana A.
; APPLICANT: Ryan, Maureen C.
; TITLE OF INVENTION: Epiligrin, an Epithelial Ligand for
; TITLE OF INVENTION: Integrins
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christensen, O'Connor, Johnson, and Kindness
; STREET: 1420 Fifth Avenue
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98101-8100
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/600,982
; FILING DATE: 02-SEP-1994

```

```

; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Shelton, Dennis K.
; REGISTRATION NUMBER: 26,997
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 682-8100
; TELEFAX: (206) 224-0779
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1994 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; DESCRIPTION: cDNA sequence corresponding to FIGURES
; DESCRIPTION: 11A-11C
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; US-08-600-982-22

Alignment Scores:
Pred. No.: 0.0915 Length: 1994
Score: 105.50 Matches: 112
Percent Similarity: 35.97% Conservative: 79
Best Local Similarity: 21.09% Mismatches: 201
Query Match: 2.84% Indels: 140
DB: 3 Gaps: 25

US-10-008-355-2 (1-712) x US-08-600-982-22 (1-1994)
Qy 69 PheGlyGlyGlyCysThrGlyIleThrValSerAspGlnGlyLeuIlePheThrAsnHis 88
Db 514 TTCGAGGTGAGTGTCCACCAATGCGATTGTACAGCAATGAGCCAGCTGGCAGCTGCAT 573
Qy 89 HisCysGlyTyrGlyAlaIleGlnSerGlnSerThrValAspHisAsp----- 104
Db 574 CCCGTGACGTGAGGCTGCATTA--AACCAAGAAACCAAGATTAACACCTGCGAGAGAA 630
Qy 105 -----TyrLeuArgAspGlyPheValSerArgThr 114
Db 631 TGTGATGATGTCGACACCTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 681
Qy 115 MetGlyGlyGluLeu-----ProIleProGlyLeuSer----- 125
Db 682 ATGGGCGAGAGCTCCGCTGTGCAAGTCTGACGTGACGAGGAGGAGGAGGAGGAGGAGG 741
Qy 126 -----Val 126
Db 742 CTTCGTGAGCAGATGAGGACATGAGACCCAGGCCAAGACCTGAGAGATTCAGTTCCTC 801
Qy 127 LysTyrLeuArgLysIleValLysValThrAspLysValGluGlyGlnLeuLysGlyIle 146
Db 802 AACCAACGTTCTGCTGATTAATCATGATCAATTAAGAAAGCCGCGAAGAGAACTG 861
Qy 147 ThrAsp-----GluMetGluArgLeuArgLysAlaGlnGluValLysGlnGluLeu 163
Db 862 ACTGATTTGAATCAAGAAATTTGAGACTTTCAGAAAGAGCTCAAGTAAATTCACAGAAA 921
Qy 164 AlaLysLysGluAsnAlaAspGluAsnGlnLeuLysIleValGluProPheTyrSerAsn 183
Db 922 GCACAAACATTAACCAACATGTTAATCGGCAACA-----CAAGCGCA 966
Qy 184 AsnGluTyrPheLeuIleValTyrAspValPheLysAspValArgMetValPheAlaPro 203
Db 967 AAAGAACTGATGAGAGATTAATAAATGATCCGCAATGTCACATCTTTAAAGACAG 1026
Qy 204 ProSerSerValGlyLysPheGlyGlyAspThrAspAsnTrpMetTrpProAlaGlnHisThr 223
Db 1027 ATCTGTGGACAGATGAGAGGAAACACAGCTGCT-----TCA 1065
Qy 224 GlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgProAlaGluTyrSer 243

```

```

Db 1066 GGTGACTTTCC-----AGAGAGTGGGCTGAAGCCACGCATGATGAGGAAGTGGCG 1119
QY 244 LysAspAsnLysProTyrLysProValTyrPheAlaIleValSerMetIleGlyTyrLys 263
Db 1120 AACAGCAAC-----TTTGAAAGCAGCTCAGAGAACAGACA 1155
QY 264 AlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrLeuThrSer 283
Db 1156 GGTGATAAAGGAGATCGCAGCTTGTGCTG-----AACGGATTAAGGAGCC 1200
QY 284 TrpGlyValGluAspArgIleGluAsnGluAsnAsnProArgIleGluValArgGlyIle 303
Db 1201 TGG-----CAGAAAACCCAGCGGGGAGAACAAAT-----GGGCTT 1236
QY 304 LysGlnGlyIleThrLys-----GluAlaMetSerAlaAspGlnAlaThr 318
Db 1237 GCTAACAGATTCGGGATTTCTTAATGATACGAAGCCAAACTCAGTACCTTCTGCT 1296
QY 319 ArgIleLysTyrAlaSerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsnSerIleGly 338
Db 1297 CGGCTGCAGGAGGACAGCTGCCAAGCCAGAGCAAGCAAT-----GGC 1338
QY 339 MetAsnArgGlyLeuAlaArg---LeuAspValIleGlyArgLysArgAlaGlu----- 355
Db 1339 TTGAAACCAAGAAAACGAGAGCTTTGGAGCCATTCAAGACAACTGAAAAGAAATTAAT 1398
QY 356 -----GluArgAlaPheAlaAspTrpIleArgLysAsnGlyLysSerAlaValTyrGly 373
Db 1399 TCCCTGCAGATGATTTCCCAAGATATCTAACCACTGCAGACTCATCTTTGTTCCAAAC 1458
QY 374 AspVal---LeuSerSerLeuGluLysAlaTyrLysGlu-----GlyAlaLys 388
Db 1459 AACATATGCCCTCAGCTGATGAGAAAACCCAGAAAGAAATATGAAAATTAAGCTGCCACT 1518
QY 389 AlaAsnArgGluMetThrTyrLeuSerGluThrLeuPheGlyGlyThrGluValValArg 408
Db 1519 TTAAATGAACAGACAGACAGACTAAGTGAACAAGTA-----AGAGAACTTTCCACA 1569
QY 409 PheAlaGlnPheAlaAsnAlaLeuAlaThrAsnProAspAlaHisAlaGlyIleLeuLys 428
Db 1570 TCTGCTGGC---AAAGCAATCCCTTGTGAGGAGGAGCAAGAAACAGCCG-----CGG 1617
QY 429 SerLeuAspAspLysTyrLysAspTyrLeuProSerLeuAspArgLysValLeuProAla 448
Db 1618 TCCCTTACAAGAG-----CTGGCAAG 1638
QY 449 MetLeuAspIleValArgArgArgIleProAlaAspLysLeu----- 462
Db 1639 CAGCTGGAAGAGATCAAGAGAAAGCCACCGGGGATGAGCTGTGCTGTGAT 1698
QY 463 ---ProAspIlePheLysAsnValIleAspLysLysPheLysGlyAspThrLysLysTyr 481
Db 1699 GCCGCACCGCCAGCAAGAAACATCCATATGCCATCAAGCGGCGAGAGACGAGCAAC 1758
QY 482 AlaAspPheValPheAspLysSerValValProTyrSerAspLysPheHisAlaMetLeu 501
Db 1759 AGGGCTCGCAGTGCATCTGAATCTGCCCT--CCAGACAGTGAATAAGGAAGATCTGCCAAG 1817
QY 502 LysSerMetAspLysGlu-----LysPheAlaLysAlaIleGlu 514
Db 1818 AAAAGTAAACCTGAGTGTCCAAACAGTATTAAGTGTAAAGAAACCAAGATGACACAA 1877
QY 515 LysAspProAlaValGluLeuSerLysSerVal 525
Db 1878 AAGAGCTAAAGCAAGAACTCAAGTCCAGCTCTC 1910

```

```

RESULT 10
PCT-US94-10261A-22
; Sequence 22, Application PC/TUS9410261A
; GENERAL INFORMATION:
; APPLICANT: Carter, William G.
; APPLICANT: Gil, Susanna A.
; APPLICANT: Ryan, Maureen C.

```

```

; TITLE OF INVENTION: Epiligrin, an Epithelial Ligand for
; TITLE OF INVENTION: Integrins
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESS: Christensen, O'Connor, Johnson, and Kindness
; STREET: 1420 Fifth Avenue
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98101-8100
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/10261A
; FILING DATE: 02-SEP-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Shelton, Dennis K.
; REGISTRATION NUMBER: 26,997
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 682-8100
; TELEFAX: (206) 224-0779
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1994 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; DESCRIPTION: cDNA sequence corresponding to FIGURES 11A-11C
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; PCT-US94-10261A-22

Alignment Scores:
Pred. No.: 0.0915 Length: 1994
Score: 105.50 Matches: 112
Percent Similarity: 35.97% Conservative: 79
Best Local Similarity: 21.09% Mismatches: 201
Query Match: 2.84% Indels: 140
DB: 5 Gaps: 25

US-10-008-355-2 (1-712) x PCT-US94-10261A-22 (1-1994)

QY 69 PheGlyGlyGlyCysThrIleThrValSerAspGlnGlyLeuIlePheThrAsnHis 88
Db 514 TTGGAGAGTAGCTGCCAACCATGCTGTGAACAGCAATGGCCAGCTGGCGAGCTGTGAT 573
QY 89 HisCysGlyTyrGlyAlaIleGlnSerGlnSerThrValAspHisAsp----- 104
Db 574 CCCCTGACTGAGGCTGCATA--AACCAAGAAACCAAGATTAACAACTCTGCAGACAA 630
QY 105 -----TyrLeuArgAspGlyPheValSerArgThr 114
Db 631 TGTGATGATTGGACACACGTGTGATGACCCCTCGAAGAC-----CTGGCCACC 681
QY 115 MetGlyGluGluLeu-----ProIleProGlyLeuSer----- 125
Db 682 ATGGGCGAGCAGACTCCGCTGCTCAAGTCTCAGCTGCAAGGGCTGAGTCCAGCCAGGG 741
QY 126 -----Val 126
Db 742 CTTTGGAGCAGATGAGCAGCATGAGAGACCCAGGCCAAGGACCTGAGGATCAGTTGCTC 801
QY 127 LysTyrLeuArgLysIleValLysValThrAspLysValGlnGlyGlnLeuLysGlyIle 146
Db 802 AACTACCGTTCGCAATTCAATCATGATGATCAAAAATAGAAAGGCTGTGAAAGAGAACTG 861
QY 147 ThrAsp-----GluMetGluArgLeuArgLysAlaGlnGluValCysGlnGluLeu 163

```

```

Db 862 ACTGATTGGAATCAAGATTTTGAGACTTTGCAAGAAAGGCTCAAGTAATTCAGAAAA 921
Qy 164 AlAluyslgluasnAlaaspgluasnGlnleucysIleValIGluProPheTySerAsn 183
Db 922 GCACAAACATTAAACAAATGTTAATCGGCAACA-----CAAAAGCGCA 966
Qy 184 AsnGluTrPheLeuIleValTyAspValPheLysAspValArgMetValPheAlaPro 203
Db 967 AAGAACTGGATGTGAAGATTAAATATGTCATCCGGAATGTGCACATCTTTTAAACAG 1026
Qy 204 ProSerSerValGlyLysPheGlyLysPheThrAspAsnTrpMetTrpProAlaGlnSer 223
Db 1027 ATCTCTGGACAGATGAGAGAGAAACAGTGCCT-----TCA 1065
Qy 224 GlyAspPheSerValPheAlaGlyValAlaLysAspAsnArgProAlaGlyTySer 243
Db 1066 GGTGACTTTTCC-----AAGAGCTGGGCTGAAACCCGACGATGAGGAACTGCGG 1119
Qy 244 LysAspAsnLysProTyLysProValTyPheAlaAlaValSerMetGlnTyTyLys 263
Db 1120 AACAGAAC-----TTTGAAAGACCTCAGAGAAAGCA 1155
Qy 264 AlaAspAspTyAlaMetTrpIleGlyPheProGlySerThrAspArgTyLysThrSer 283
Db 1156 GCTGATTAAGAGGAGTGCAGCTCTTGCTG-----AACCGATTAAGAGACC 1200
Qy 284 TrpGlyValGluAspArgIleGluAsnGluAsnAsnProArgIleGluValArgGlyIle 303
Db 1201 TGG-----CAGAAACCCACACAGGGGAGAACAAAT-----GGGCTT 1236
Qy 304 LysGlnGlyIleTrpLys-----GluAlaMetSerAlaAspGlnAlaThr 318
Db 1237 GCATACAGTATCCGGGCTTCTTAATGAATACGAAAGCAAACTCAGTACGCTCGTCT 1296
Qy 319 ArgIleLysTyAlaSerLysTyAlaGlnSerAlaAsnTyTrpLysAsnSerIleGly 338
Db 1297 CGGCTGCAGAGGAGCGTGCACCAAGCAAGCAGCAAT-----GGC 1338
Qy 339 MetAsnArgGlyLeuAlaArg--LeuAspValIleGlyArgLysArgAlaGlu----- 355
Db 1339 TTGAACCAAGAAACGAGAGCTTTGGAGCCATTCAGACAAAGTAAATTAAT 1398
Qy 356 -----GluArgAlaPheAlaAspTrpIleArgLysAsnGlyLysSerAlaValTyGly 373
Db 1399 TCCTTCAGAGATGATTTCCACCAAGTATCTAACCACTGACATCTTTGTTCAACC 1458
Qy 374 AspVal--LeuSerSerLeuGluLysAlaTyLysGlu-----GlyAlaLys 388
Db 1459 AACATTGCGCTGCAGCTGATGAGAAAGCAAGAGATATGAAATTAAGCTGCCAGT 1518
Qy 389 AlaAsnArgGluMetTrpTrpLeuSerGluTrpLeuPheGlyGlyThrGluValValArg 408
Db 1519 TTAATATACAGACAAAGACTTAAGTACAAAGTA-----AGAGAACTTTCAGA 1569
Qy 409 PheAlaGlnPheAlaAsnAlaLeuAlaThrAsnProAspAlaHisAlaGlyIleLeuLys 428
Db 1570 TCTGCTGGC---AAACATCTCTTGTGAGAGAGCAGAAAGACGCGG-----CGG 1617
Qy 429 SerLeuAspAspLysTyLysAspTyLysProSerLeuAspArgLysValLeuProAla 448
Db 1618 TCCTTACAGAG-----CTGGCAAG 1638
Qy 449 MetLeuAspIleValAlaGlyArgGlyIleProAlaAspLysLeu-----462
Db 1639 CAGCTGGAGATGAGAAAGCAAGCCAGCGGATGAGTGTGCGCTGTGTGAT 1698
Qy 463 ---ProAspIlePheLysAsnValIleAspLysLysPheLysGlyAspTrpIleTyLysTy 481
Db 1699 GCCGCGCACCGGCTACGAAACATCTCAATGCCATCAAGGGCGGAGACCGGCAAC 1758
Qy 482 AlaAspPheValPheAspLysSerValValProTySerAspLysPheHisAlaMetLeu 501

```

```

Db 1759 AGGCTCGCAGTGCATCTGAATCTGCCCT--CCAGACAGTATTAAGAGATCTGCCAG 1817
Qy 502 LysSerMetAspLysGlu-----LysPheAlaLysAlaIleGlu 514
Db 1818 AAAAGCTAAACCCCTGAGTTCCAACTGATTAACCTTTAAAGAACCAAGATGACACAA 1877
Qy 515 LysAspProAlaValGluLeuSerLysSerVal 525
Db 1878 AAGAGCTAAAGCAAGAGATGATCAGCTC 1910

RESULT 11
US-08-811-583-1
; Sequence 1, Application US/08811583
; Patent No. 6218142
; GENERAL INFORMATION:
; APPLICANT: Wassenecker, Michael
; APPLICANT: Riedel, Leonhard
; APPLICANT: Schiebel, Winfried
; APPLICANT: Sanger, Heinz
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
; TITLE OF INVENTION: POLYPEPTIDES HAVING THE ENZYMATIC ACTIVITY OF AN
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH & NEAVE
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/811,583
; FILING DATE: 05-MAR-1997
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley, James F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: MPG-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-596-9000
; TELEFAX: 212-596-9090
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3731 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Tomato
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 194..3535
; US-08-811-583-1

Alignment Scores:
Pred. No.: 0.354 Length: 3731
Score: 104.00 Matches: 133
Percent Similarity: 32.06% Conservative: 77
Best Local Similarity: 20.31% Mismatches: 232
Query Match: 2.80% Indels: 213
DB: 4 Gaps: 31

US-10-008-355-2 (1-712) x US-08-811-583-1 (1-3731)
Qy 92 TyGlyValAlaIleGlnSerGlnSerThrValAspHisAspTyLysLeuArgAspGlyPheVal 111
Db 1859 TATGCTGATATAAAGATGTTGCGGTGTGATCCGATTCATCAATGAAGATGTGTCTTG 1918

```

```

QY 112 SerArgThrMetCylGlu-----GluLeuProIleProGlyLeuSer 125
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1919 AGAAGAGCATGTGCAAAATATGAAATCAGACAAACATAAATGATGCTTGGATGAGC 1978
QY 126 ValLysTyr-----LeuArgLysIleValLys 134
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1979 ---AAATATCAGCCTGTATCTTAATCTCAACTGATTAACGCTCTTGTCTACACTTGA 2035
QY 135 ValThrAspLysVal---GluGlyGlnLeuLysGlyIleThrAspGluMetGluArgLeu 153
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2036 GTGAAGATGAAAGTTCTCAGACAGAGCAAAAGAAAGAACTGTGATCAGCTTGATGCTATC 2095
QY 154 -----ArgLysAlaGlnGluValCysGlnGlnLeuAlaLysLysGlnAla 169
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2096 TTGCATGATCTTTTGAAGGCACAGAGGCTTTGGAATGATGTCCTCGAGAGAACT 2155
QY 170 AspGluAsnGlnLeuCysIle-----ValGluProPheTyrSer 182
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2156 AATATTCTCAAGGCAATGCTAAAGCTGGTTATAGCCGTAGCTGAGCCCTTTCTTTCA 2215
QY 183 AsnAsnGluTyrPheLeuIleValTyrAspValPheLys-----AspVal 197
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2216 ---ATGATGTTCGAACACTTCGCGCATCCAAAGTTGCTCGATTGG 2257
QY 198 Arg-----MetValPheAlaProPheSerSerVal----- 207
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2258 CGGACTAGATCAAGATATTTATTTCCAAATGGAAGAAACATGAGTGGATTGGATGA 2317
QY 208 -----GlyLys 209
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2318 TCCAGAACCTTGGAAATATGCTAGCTGTTGTTCACTTACTGCTGACATGAGAGAG 2377
QY 210 PheGlyGlyAspThrAspAsnTrpMetCysTrpProArgHisThrGlyAspPheSerValPhe 229
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2378 TTTTCGAGAGATTACATCATTTATTAACAGAGATCCACCAAGATTAATTCATTCTG 2437
QY 230 ArgValTyrAlaGlyAlaAspAsnArgProAlaGluTyrSerLysAspAsnLysProTyr 249
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2438 AAGGGAATATGCTGTTGCCAAAATAATCCATGCTGATCCTGATGATATGCTGTTTCA 2497
QY 250 LysProValTyrPheAlaIleValIleSerMetGlnGlyTyrLysAlaAspPyrAlaMet 269
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2498 AAGGCTGTAAATGCTTCGACGCTGCAC-----CACATGGTAGAT----- 2536
QY 270 ThrIleGlyPhePro-----GlySerThr 277
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2537 TGTGTGTATTCCTTCAGAAAGAAAGAAACCTCATCCGAATGAATGTTCTGGAGTGTAT 2596
QY 278 -----AspArgTyrLeuThrSerTrpGlyValGluAspArgIleGluAsnGluAsn 294
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2597 TTGGATGGGATATCTACTTTGTTGCTG-----GAT 2629
QY 295 AsnProArgIleGluValArgGlyIleLysGlnGlyIleTrpLysGluAlaMetSer--- 313
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2630 CAAGACATGATCCCGCCAGGCAAGTCAGCCGATGAATATCTCCACACCCACACATTA 2689
QY 314 ---AlaAspGlnAlaThrArgIleLysTyrAlaSerLysTyrAlaGlnSerAlaAsnTyr 332
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2690 CAGTTGGACCATGATGTCAACATTTGAGAAAGTTGAAGAGATAC-----TTCACCACTAT 2743
QY 333 TrpLysAsn---SerIleGlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArg 351
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2744 ATTGTGAATGACAGTTTGGGAATC-----ATAGCAAAAGCCCATGTGATTTTGA 2794
QY 352 LysArgAlaGlnGluArgAlaPheAlaAspTrpIleArgLysAsnGly----- 367
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2795 GACAGAGAACTGATATGTCATGATGATGCATGCAGCAAAAACCTGCTAGACTTTTCA 2854
QY 368 -----LysSerAlaValTyrGlyAspValLeuSerSerLeuGlu--- 380
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2855 ATTGCAGTGACTTTCCAAAGAGCTGCTGTTCCCGCTGAATAATACATCATGATTGGCCCT 2914

```

```

QY 381 LysAlaTyrLysGlnGluValAlaLysAlaAsnArgIleMetThrTyrLeuSerGluThrLeu 400
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2915 AAGAGATACCACACTTCATGTGATGAAGCGGAGACAGACAGCTATATATCCAAAGAGTT 2974
QY 401 PheGlyGlyThrGluValValArgPheAlaGlnPheAlaAsnAlaLeuAlaThrAsnPro 420
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2975 ATTTGA----- 2980
QY 421 AspAlaHisAlaGlyIleLeuLysSerLeuAspLysTyrLysAspTyrLeuProSer 440
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2981 -----AAGCTTTTCAGGAAAGTGAAGACAAACACTCTGAG 3016
QY 441 LeuAspArgLysValLeuProAlaMetLeuAspIleValArgArgIleProAlaAsp 460
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3017 GCTAGC---TCTATCGCGACTTCACAAGAGATGTTGCAAGAGATCATATGATGCTGAT 3073
QY 461 LysLeuProAspPheLysAsnValIleAspLysLysPheGlyAspThrLysLys 480
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3074 ATGGAAGTTGATGATTTGATGATTAACATTAATGACGAAAGCTTTT---GACTACAAAAC 3127
QY 481 TyrAlaAspPheValPheAspLysSerValIleProTyrSerAspLysPheHisAlaMet 500
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3128 GATATATGACAAACAGCTGGCTAATTAATGACACTACTATAGCATATAAACAAGCCTGGA 3187
QY 501 Leu-----LysSerMetAspLysGluLysPheAlaLys 511
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3188 ATACTTACTGTGCGCATTTATGAAGCATCAAAAACCTTTTGACCGCAGAAAGATGCTGAG 3247
QY 512 AlaIleGluLysAspProAlaValGluLeuSerLysSerValIleAlaIleAlaIleArgAla 531
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3248 GCCATTAGT-----GTTGCTGTGAGGCC 3271
QY 532 IleGlnAlaAspAlaMetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPhePhe 551
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3272 TTGAGAGAGAGGCAAGA----- 3289
QY 552 AlaGlyLeuArgGluMetCysTrpGlyArgAlaLeuProSerAspAlaAsnPheThrMet 571
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3290 -----GCCGTGTTCAAGAG 3304
QY 572 ArgMetSerTyrGlySerIleLysGlyGluProGlnAspGlyValArgPyrAsn--- 590
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3305 CGT-----AATGATATATGATGATCATGTTATCCAAAGCCTTCGGCTTGATCCAGCTT 3355
QY 591 ---TyrHisThrThrGlyLysGlyVal-----LeuGlnLysGlnAspProLys 605
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3356 ACATATCATCTCAACATATTTGGGTTGCTACATCAAGGGTTGAAAAGACCTCATTTCAAT 3415
QY 606 SerAspGluPheAlaValGlnGluAsnIleLeuAsp---LeuPheArgThrLysAsnTyrGly 625
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3416 ACCTTTCCCTGGTGTGTTATGACCGATTAATCCAGATTAAGAGACAAACACAGTA-- 3473
QY 625 ValArgTyrAlaGluAsnGlyGlnLeuHisIleAlaPheLeuSer 639
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3474 -----ACAGGCCAGTTCTCAACTTGTCACTCTCA 3503

RESULT 12
US-09-514-302-1
; Sequence 1, Application US/09514302
; Patent No. 638959
; GENERAL INFORMATION:
; APPLICANT: HATADA, Yuji
; APPLICANT: IGARASHI, Kazuaki
; APPLICANT: OZAKI, Katsuya
; APPLICANT: KAWAI, Shuji
; APPLICANT: ARA, Katsutoshi
; APPLICANT: ITO, Susumu
; TITLE OF INVENTION: GENE FOR ENZYME HAVING BOTH ALKALINE PHOSPHATASE AND
; FILE REFERENCE: 2173-105P
; CURRENT APPLICATION NUMBER: US/09/514,302
; CURRENT FILING DATE: 2000-02-28
; EARLIER APPLICATION NUMBER: 08/952,084

```

EARLIER FILING DATE: 1997-11-10
 NUMBER OF SEQ ID NOS: 14
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO 1
 LENGTH: 6142
 TYPE: DNA
 ORGANISM: Bacillus sp.
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (145)..(5958)
 FEATURE:
 NAME/KEY: sig_peptide
 LOCATION: (145)..(240)
 FEATURE:
 NAME/KEY: mat_peptide
 LOCATION: (241)..(5958)
 US-09-514-302-1

Alignment Scores:

Pred. No.:	1-12	Length:	6142
Score:	102.50	Matches:	108
Percent Similarity:	31.46%	Conservative:	71
Best Local Similarity:	18.98%	Mismatches:	199
Query Match:	2.76%	Indels:	191
DB:	4	Gaps:	25

US-10-008-355-2 (1-712) x US-09-514-302-1 (1-6142)

QY 231 VALTYRAGLYALAAspAsnArgProAla-----GluTyrSerLysAspAsn 246
 DB 1036 GTGACGGCAGGTGTGAAAACCTTGCTATCATCGCATGTGATGAAATTCGAAATGCCAT 1095
 QY 247 LysProTyrLysProValTyrPheAlaAlaValSerMetGlnGlyTyrLysAlaAsp--- 265
 DB 1096 GAGGAGATCTGACCTTACAAAGTTCCAGCGCTACTATTACAGCTGAAAGAGATTTC 1155
 QY 266 AspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrLeuThrSerTyrGly 285
 DB 1156 GACTGGGATCATCTGTGGTTATTATTATGCTGACGATCATCTTT----- 1203
 QY 286 ValGluAspArgIleGluAsnGluAsnAsnProArg----- 297
 DB 1204 -----GATGGGAGATCATCGAACAATGACCCCTCATGATGCGTATGACACACAGCAG 1257
 QY 298 -----IleGluValArgGlyIleLysGln----- 305
 DB 1258 TCTGTACATACCAAGCGGAGATTTTAAGGAGATCAACGAAAGCGTTGATTACTTGAC 1317
 QY 306 -----GlyIleTyrLysGluAlaMetSerAlaAspGlnAlaThrArgIle 320
 DB 1318 GAGCTTGAATCAATACATTCGATCAGTCCGGTT-----GTCCATTAATATC 1365
 QY 321 LysTyrAlaSerLysTyrAlaGlnSerAlaAsn-----Tyr 332
 DB 1366 AATATTGATGTTCGACACAGCAGTGAAGACGATACACATATTATGCTTACACAGCGCTAT 1425
 QY 333 TrpLysAsnSerIleGly----- 338
 DB 1426 TGGGCGGATTAATTCGGGGAATTAACCCGCAATTCGTTCCATGGCGGATTTCCATGA 1485
 QY 339 -----MetAsnArgGlyLeu-----AlaArgLeuAspValIle 349
 DB 1486 ATGATTGATGGCGGACATCAACGCGCATTTAAATCATGTTGATGTGTGTTGAATCAC 1545
 QY 350 -----GlyArgLysArgAla-----Glu 355
 DB 1546 ACTGGTTATGATGAACACAGGTGACACAGCAGTGTGCGAACTTCCGACAGATGAG 1605
 QY 356 GluArgAla---PheAlaAspTrpIleArgLysAsnGlyLysSerAlaValTyrGlyAsp 374
 DB 1606 GACCGAGCTGCTTTCACGGAATGCTTCGTGATGCGGAGATCTGCGAAGTTCGAGCGGAG 1665
 QY 375 ValLeuSerSerLeuGluLysAlaTyrLysGluGlyAlaLysAlaAsnArgLysMetThr 394

DB 1666 ----CTTGGCGCTTCCAGATTTTTCGACGGAACCCGAGTCCGTGACAGCGTGTG 1722
 QY 395 Tyr-----LeuSerGluThrLeuPheGlyGlyThr----- 404
 DB 1723 CATGCGACAGCGACTGATCGAAAAGTCCAGGACGCAAGCGCAACCATCGATTAT 1782
 QY 405 -----GluValValArgPheAlaGlnPheAlaAsnAla 415
 DB 1783 TTCCGTGTGACACCCGTCAAGCATGTGGAAGACACCACCTTGATGGCTTAAAAATGCT 1842
 QY 416 LeuAlaThrAsnProAspAlaHisAlaGlyIleLeuLysSerLeuAspLysTyrLys 435
 DB 1843 TTGACAAAAGCCGATGCCGGAACACAAAGCTGATTGGGGAAGCATGGGAGACCAATGCAAT 1902
 QY 436 AspTyrLeuProSerLeuAspArgLysValLeuProAlaMetLeuAspIleValArgArg 455
 DB 1903 GACGACCTGATTATCTGAACAGCGGATGATGATTTCTTACTGATTTT----- 1953
 QY 456 ArgIleProAlaAspLysLeuProAspIlePheLysAsnValIleAspLysLysPheLys 475
 DB 1953 ----- 1953
 QY 476 GlyAspThrLysTyrAlaValPheValPheAspLysSerValValProTyrSerAsp 495
 DB 1954 ----GATTCAAAAATTATATGCGCTGACTTGCACAAAGCAGCATGCGTTCAGCAA 2010
 QY 496 LysPheHisAla-----MetLeuLys 502
 DB 2011 AAACCTGAGCGCGGTACAGCAAGTTGACAAATATCTGCAACACTTGTCATAATTATAGGA 2070
 QY 503 SerMetAspLysGluLysPheAlaLysAlaIleGluLysAspProAlaValGluLeuSer 522
 DB 2071 AGCCATGACGAAGACCGCTTATGATGAGTGTGGAAGGAGAC-----CTTGGC 2118
 QY 523 LysSerValIleAlaAlaIleAlaArgAlaIleGlnAlaAspAlaMetAlaAsnAlaTyrAla 542
 DB 2119 AAGTATCAAGTTGCTGATCCCTTCACTGACGCGCA----- 2154
 QY 543 IleGluLysGlyLysArgLeuPhePheAlaGlyLeuAlaGlyLysMetTyrProGluArgAla 562
 DB 2155 ----AAGGTCACGCTTATCTATATACGGAAGACGCTGGCTGCTGTAG---- 2205
 QY 563 LeuProSerAspAlaAsnPheThrMetArgMetSerTyrGlySerIleLysGlyTyrGlu 582
 DB 2206 ----AAGCAT----- 2211
 QY 583 ProGlnAspGlyAlaLeuTyrLysTyrHisThrThrGlyLysGlyValLeuGluLysGln 602
 DB 2212 -----TATCCGTTATACGAACCCGACAGAAC-----ATGCCCTGGGAT 2250
 QY 603 AspProLysSerAspGluPheAlaValGlnGluAsnIleLeuAspLeuPheArgThrLys 622
 DB 2251 GATGTGATGATGAATGAATTTCTAGAGCATTTATCAAAAATTTCTAGCATTTCCGTAATGAT 2310
 QY 623 AsnTyrGlyArgTyrAlaGluAsnGlyGlnLeuHisIleAlaPheLeuSerAsnAsnArg 642
 DB 2311 AATCGAACAATTTCTTAAGGAGACCCGAAAGAGTGAAG-----GGATGTGAC 2361
 QY 643 IleThrGlyLysAsnSerLysSerProValPheAspLysAsnIlyArgLeuIleGlyLeu 662
 DB 2362 AGTGAAGATATCTTTATTTTCACGAGGTACGGGGAATAATTCGCTTATATAGAGTTTG 2421
 QY 663 AlaPheAspGlyAsnTrpAlaLysMetSerGlyAspIleGluPheGluProAspLeuGln 682
 DB 2422 -----AATACGGAAGCTGCTCGAAAGAGATGAACCTTGAACCTTGGTTCTTCA 2469
 QY 683 ArgThrIleSerValAspIleArgTyr 691
 DB 2470 GAAGCAGTGTGACGAGC---CGCTAT 2493
 RESULT 13
 US-08-816-105A-2


```

: Sequence 2, Application US/08816105A
: Patent No. 5989882
: GENERAL INFORMATION:
: APPLICANT: Crueger, Anneliese; Dellwieg, Hans-Georg; Lenz,
: APPLICANT: Jurgens, Schroder, Werner; Pape, Hermann;
: APPLICANT: Goeke, Klaus; Schaper, Beate; Hemker, Michael;
: APPLICANT: Piepersberg, Wolfgang; Distler, Jürgen;
: APPLICANT: Stralman, Ansgar
: TITLE OF INVENTION: PROCESSES FOR PREPARING ACARBIOSYL
: TITLE OF INVENTION: TRANSFERASE AND FOR USING IT IN THE
: TITLE OF INVENTION: CONVERSION OF ACARBOSE HOMOLOGUES
: TITLE OF INVENTION: INFO ACARBOSE, FOR THE PREPARATION
: TITLE OF INVENTION: OF ACARBOSE HOMOLOGUES
: NUMBER OF SEQUENCES: 16
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Sprung Kramer Schaefer & Briscoe
: STREET: 660 White Plains Road
: CITY: Tarrytown
: STATE: New York
: COUNTRY: USA
: ZIP: 10591-5144
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
: COMPUTER: Apple Macintosh 6500
: OPERATING SYSTEM: System 7.5
: SOFTWARE: WordPerfect 3.5
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/816, 105A
: FILING DATE: 14-MAR-1997
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: DE 19625269.5
: FILING DATE: 25-JUN-1996
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: DE 19611252.4
: FILING DATE: 22-MAR-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Kurt G. Briscoe
: REGISTRATION NUMBER: 33,141
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (914) 332-1700
: TELEFAX: (914) 332-1844
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2582 nucleotides
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
:
: US-08-816-105A-2
:
: Alignment Scores:
: Pred. No.: 0.332 Length: 2582
: Score: 102.00 Matches: 108
: Percent Similarity: 30.33% Conservative: 67
: Best Local Similarity: 18.72% Mismatches: 175
: Query Match: 2.74% Indels: 227
: Gaps: 33
:
: US-10-008-355-2 (1-712) x US-08-816-105A-2 (1-2582)
:
: Oy 62 TlealasnalaValalIlePheglYglYcYsthrGlyIlethrhValSeraspIn 81
: Db 164 CTTCCGCTGCTGCCGCGCTACGCGCGTCATCG----- 199
:
: Oy 82 GlyleuIlePhePthrasnHnIsHscYsgLYrGlyAlaIle-----Gln 96
: Db 200 -----CGCGCGGCGCTTGGCCGCTGTCGCGGTACCGCGGTAAACCGCGCCAG 244
:
: Oy 97 SerGlnSer-ThrValaspHIsAsPtyrLeuIdgAspGlyPheValSerAgtThmMetGl 116
: Db 245 TCTCAAGCAGGACCTCTGCTACCAAGATCGCACCGACCGGTTCAAGCAGGAGCCCGCGC 304

```

```

Oy 116 YglGlnIleuPro-----IleProGlyLeuSerValIlystYrIleuArgIysIleVally 134
Db 305 GAACACAAATCCGGGCAACGATGCCCGGCATG-----TTGCCGCAACAGACCA 352
Oy 134 SvalThrAspIysValGlu-----GlyGlnIleuYsgIlyIlethras 148
Db 353 GCTAAGACCGGCGGAGAGAGTGGCTCAATATCATGTGGAGGTACTTCCCGGCATCACCA 412
Oy 148 pGluMetGluArgIleuArgIysAlaGlnIleValcYsgInGlnIleuAlaYsIysGluas 168
Db 413 GCGATGAGTCACTCAAG-----AA 433
Oy 168 nAlaspGluasnInIleuYcYsIleValGluProPheYrSerasnInGlyrPhele 188
Db 434 CTTGGCGTCCGCGCGATCTGGATC-----TCGCCCGACGTGCACAC----- 476
Oy 188 uIleValIyraspValPheYsAspValArgMetValPheAlaProProSerValGl 208
Db 477 -----ATCAAGCTCCGGCGAAGCGGCCAC 502
Oy 208 YLysPheglYglYaspThraspAsnTrpMetTrpProArg----- 221
Db 503 CGGTACACAGC-----TACTGGCGCGGCACTTCAAGCGCTCGAAGA 547
Oy 222 -----HisThrGlyaspH 226
Db 548 GCACCTTCGACCGACGAGAGATTGACCGCTGTGCGCGCGGCGACCGCACCAT 607
Oy 226 eSerValPheArgValYrAlaGlyAlaAspAsnArgProAlaGluYrSerIysAspAs 246
Db 608 CAAGTGATCATGGACGTGACGCGCAACGCGCAACCCCGGACCAAGCGCGAGAGCGG 667
Oy 246 nLYsProTYrLYsProVALyrPheAlaAlaValSerMetGlnLYrLYsAlaAspAs 266
Db 668 C-----GCCCTCTACGACGATGGCGACGTGTCGCGAGTACGGCGGAGCAG 715
Oy 266 P-----TyrAlaMetThrIleGlyPheProGlySerThrAspArgTYrLe 281
Db 716 TGCCGGGCACTTCCACCAAGCGCGCGGATCGG-----GACTTCAACGATCTTCCA 769
Oy 281 uThrSer---TrpGlyValGluAspArgIleGlu---AsnGluasnProArgIleGl 299
Db 770 GCACCACTACTACAGCCTCGCGGCACATCGCCGACCTGCACGAGACCGCGGGGTC-- 827
Oy 299 uValArgIlyIleYsgInGlyIleTrpYsgIuAlaMetSerAlaAspGlnAlaThrAr 319
Db 828 -----GACGACGCTGCTCAA 841
Oy 319 GlleYsTYrAlaSerIyTYrAlaGlnSerAlaAsnTYrTrpLYsAsnSerIleGlyMe 339
Db 842 G-----GACGACCGCAACTACTGG-----AT 862
Oy 339 tAsnArgIlyLeu-----AlaArgLeuAspValIleGlyrArgYsArgAlaGlu---Gl 356
Db 863 GGACCGGGGGTTCACGCGCATCCGGGTGACGCGGTCAAGACATGCCGCTGACCTGGCA 922
Oy 356 uArgAlaPheAlaAspTrpIleArgIysAsnGlyLYsSerAlaValYrGlyAspValle 376
Db 923 GCGGTCTCTCCGCGACGCGGTGCACCTCGCACAAAGCGCGGCATCTTCGCGGAG----- 977
Oy 376 uSerSerLeuGluIuYsAlaTYrLYsGluGlyAlaLYsAlaAsnArgIuMetThrTYrLe 396
Db 978 -----TGTTACATGGCGGACCATGTC-----GATCCGCTCTACGC 1012
Oy 396 uSerGluThrIleuPheglYglYthrGluValValArgPheAlaGlnPhe---AlaAsnAl 415
Db 1013 CGACCAAGTCAAGTTCGCCAACACACCGCGCATCGCGGCATGACTTCAACCAACCG 1072
Oy 415 aLeuAlaThrAsnProAspAlaHIsAlaGlyIleuYsSerLeuAspAspIysTYrLY 435
Db 1073 CTCGATCCGCGACACACTTTCGCGGCGCGGCTCGATGATGAAGTCCCTGGAC----- 1121
Oy 435 sAspTYrIleuProSerLeuAspArgLYsValIleuProAlaMetLeuAspIleValArgAr 455

```


Db 1121 ----- 1121
QY 455 garGIIeProAlaAspLysLeuProAspIlePheLysAsnValIleAspLysLysPheLys 475
Db 1122 -----GC 1123
QY 475 sglYAspThrLysLysTyraLAspPheValIleAspLysSerValAlProTyrSerAs 495
Db 1124 GGGATACACCAAGACCAACCGGACCTACCTACGACAGACAGATGTGATCAGTTCCGGA 1183
QY 495 pLysPheHisAlaMetLeuLysSerMetAspLysGluLysPheAlaLysAlaIleGluLys 515
Db 1184 C-----AACAGGACACCGCGGCTTGGC---ACGCTCAACAG 1219
QY 515 sAsProAlaValGluLeuSerLysSerValIleAlaAlaAlaArgAlaIleGlnAlaAs 535
Db 1220 CGATCCGGCG-----GCCCTGACCGGCGCTC----- 1247
QY 535 palMetAlaAsnAlaTyraLAlaIle---GluLysGlyLysArgLeuPheAlaGlyLe 554
Db 1248 -----GCCCTTCCTGCTCACACCGCGGCTTGGCCTTGTCTACAGCGC-- 1292
QY 554 uATgGluMetTyrProGlyArgAlaLeuProSerAspAlaAsnPheThrMetArgMetse 574
Db 1293 -ACCGAGCACTAC-----CTGCACAAAGCACAC-----GG 1321
QY 574 rTyrgLysSerLysGlyTyrGluProGlnAspGlyAlaTrpTyrAsn 590
Db 1322 TGAGGCGACGACACAGGCGAAGGACCG-----TACAAAC 1355
RESULT 14
US-09-008-172-1
; Sequence 1, Application US/09008172
; Patent No. 6127602
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starch and
; FILE REFERENCE: 0358D
; CURRENT APPLICATION NUMBER: US/09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 4310
; TYPE: DNA
; ORGANISM: Streptococcus mutans
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (5)...(4295)
US-09-008-172-1
Alignment Scores:
Pred. No.: 0.736 Length: 4310
Score: 102.00 Matches: 134
Percent Similarity: 34.30% Conservative: 129
Best Local Similarity: 18.67% Mismatches: 280
Query Match: 2.74% Indels: 263
Gaps: 41
US-10-008-355-2 (1-712) x US-09-008-172-1 (1-4310)
QY 20 gLyValAlaLysAlaAspLysGlyMetTrpLeuLysAsnGluLeuAsn----- 35
Db 1463 GGTGTGATATAAGAGTGAAGAAAT-----GCCATTAATATCTTTCATTTTGAAGACT 1516
QY 36 -----GlnGluAsnLeuAspArgMetArgGluLeuGlyPheThrLeu 49
Db 1517 TGCGATGATATATGATCCCAATACATTAAGATACATAAG-----GGTGCACAAATTA 1567

QY 50 ProLeuAspSer-----LeuTyrSerPheAspLysPro----- 60
Db 1568 CCAGTATGATATATAACTGCCCTATCGCTTTATATGCTTGTGACGCGCTCTTGAAGAAA 1627
QY 61 SerIleAlaAsnAlaValIlePheGlyGlyGlyCysThrGlyLeuThrValSerAsp 80
Db 1628 GATGCAGCAATATAAATGAATTCGACGGGACTTGAAGCTGTGATACAAATATAGC--- 1684
QY 81 GlnGlyLeuIlePheThrAsnHisCysGlyTyrGlyAlaIleGlnSerGlnSerThr 100
Db 1685 -----TGATATAACCGTTGACGCAAGCTAAA---AATAGTGAACGTATG 1726
QY 101 ValAspHisAspTyrLeuArg-----AspGlyPheValSerArgThrMetGlyGluGlu 118
Db 1727 GCTAATCATATATTTTATCCCGCTCACGACAGATGCAACGATTCGCAACGTTTAA 1774
QY 119 LeuProIleProGlyLeuSerValLysTyrLeuArgLysIleValLysValThrAspLys 138
Db 1775 -----ATTGCTTAAATTTTAAAGCT-----CAG 1798
QY 139 ValGluGlyGluLeuLysGlyIleThrAspGluMetGluArgLeuArgLysAlaGlnGlu 158
Db 1799 ATTAATCCCAAAACAGATGGTTGACCTTACTTGTGATGAATGAAGCAACGCTTAA 1858
QY 159 ValCysGlnGlu-----LeuAlaLysLysGluAsnAlaAspGluAsnGlnLeuCys 175
Db 1859 ATCTACAAATAGACATGCGTCAAGGCTAAGAAAAGATACACACAAATCATATTCGACA 1918
QY 176 IleValGluProPheTyrSerAsnAsnGluTyrPheLeuIleValTyr-----AspVal 193
Db 1919 GCCATGCTTTGATGCTGCTCCATTAAGATTTCATACAGTGTTTTATGTGGATATG 1978
QY 194 PheLysAspValArgMetValPheAlaProSer----- 205
Db 1979 TACAGTATGACGCTCAATATATGCGCATTAATCCCTTATATGATGCTATGATACT 2038
QY 206 -----SerValGlyLysPhe 210
Db 2039 TTATTAAGCAGCATTAATATATGCCGCGGTGTCAGACATGAAGATCACTATGTT 2098
QY 211 GLyLysAspThrAspAsnTrpMetTrpProArgHisThrLysAspPheSerValPheArg 230
Db 2099 GAAGGTATATAAGTCAATATGATGGAT---TATACAGCGGTTTGACTGTGCTGCT 2155
QY 231 ValTyrAlaGlyAlaAsp----- 236
Db 2156 TATGTTACAGAGCTAATGAAGCTACAGATCAAGGACGAGCAACTAAACACAAGA 2215
QY 237 -----AsnArgProAla----- 240
Db 2216 ATGGCTGTCAATTACACGACAAATTAACCTTAATTAATGAATCAAAATGATTAAT 2275
QY 241 -----GluTyrSerLysAspAsnLysProTyrLysProValTyrPheAlaAlaVal 257
Db 2276 GTCAATATGGGGGTGGCGCATTAATAAATCAAGAGTACGCTCGCTTAACAACATAA 2335
QY 258 ---SerMetGlnGlyTyrLysAlaAspAspTyrAlaMetThrIleGlyPheProGlySer 276
Db 2336 GATGGTTTGCAGAGCTACACTTGTGATGCGCGCTCAATCCCTT---TATCGCAAAAGC 2392
QY 277 ThrAspArgTyrLeuThrSerTrpGlyValGluAspArgIleGluAsnGluAsnAspPro 296
Db 2393 AATGATATAAGAGATTAATGCTTTGATGCTAGTGC---ATTCAAGGTTACTGGAATCCG 2449
QY 297 ArgIleGluValArgGlyIleLysGlnGlyIleTyrLysGluAlaMetSerAlaAsp--- 315
Db 2450 -----CAAGTATACAGTTAT---TATGCGGTTTGGTTCCAGTAGAGCACTAGATTAAT 2500
QY 316 GlnAlaThrArgIleLysTyrAlaSerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsn 335
Db 2501 CAAGATGTGCTGTAGACAGCAAGCAATTAAGCAATCTACTGTCAAGCTACGATCA 2560
QY 336 SerIleGlyMetAsn-----ArgGlyLeuAlaArgLeu----- 346

```

Db 2561 TCACGCTCTTGCATTCACAAATGATTTAGACGAGTTTCTCAACTTCCACAGATTTTGTA 2620
OY 347 -----AspValIleGlyArgLysArgAla----- 354
Db 2621 ACGAAAGATTACAGACTATATCAATAAGAAAGATTGCTCAAAATGTCACACTCTTCAATATC 2680
OY 355 -----GluGluArgAlaPhe 359
Db 2681 TGGGGTGCATCTTCCCTTGAATGGACGCAATATGCTCTTCGAGAGATGGTCTTTT 2740
OY 360 AlaAspTPPIleArgLysAsnGly----- 367
Db 2741 CTAGACTCTATTATTCAAAATGTTATGCTTTGAGAGATGTTATGATCTTCTATGACT 2800
OY 368 LysSerAlaValTyrGly-----AspValLeuSerSerLeuGluLysAlaTyrLys 384
Db 2801 AAGAAATACAAATACGGTTCTCAGCAAGACATGATTAAATGCGTTAAAGCTCTGCATTA 2860
OY 385 GluGlyAlaLysAlaAsnArgLysGluMetThrTyrLeuSerGluThrLeuPheGly----- 402
Db 2861 ACGGGTATTCAGCTTATTCGGAT-----TGGGTACCAAGATCAATCTATATCTCCG 2914
OY 403 GlyThrGluValValArgPheAlaGlnPheAlaAsnAlaLeuAlaThrAsnProAspAla 422
Db 2915 GGCAAAGAAAGTGTGTA-----ACGGCTACACGCTGTCACAGAT 2950
OY 423 His-AlaGlyIleLeuLysSerLeuAspLysTyrLysAspTyrLeuProSerLeuAs 442
Db 2951 TATGGTAGATTCGCAAGACCTCTGAATCAAAATACCTGTATGCTGCCACA----- 3005
OY 442 PArgLysValLeuProAlaMetLeuAspIleValArgArgGlyIleProAlaAspLysLe 462
Db 3006 -----CTAAGAGTAAATGTAAGATTTATCAAGCAAGTATG----- 3041
OY 462 uProAspIlePheLysAsnValIleAspLysLysPheLysGlyAspThrLysLysTyrAl 482
Db 3041 ----- 3041
OY 482 aaSPheValPheAspLysSerValValProTyrSerAspLysPheHisAlaMetLeuLys 502
Db 3042 -----GCGGTCTTCCCTTAGTAAGACCGCTGCTAAGTACCTTACT 3082
OY 502 sSerMetAspLysGluLysPheAlaLysAlaIleGluLysAspProAlaValGluLeuSe 522
Db 3083 ATCTTAAACCCGACCAAAATTTCAATAGTGAAG-AAAGATTGATCCAAAGCAAAAA----- 3136
OY 522 rLysSerValIleAlaAlaAlaArgAla----- 531
Db 3137 -----ATCACAGCATGGAAAGCAAAATACTTCAATGGGACAAATATTCTAGGCCG 3186
OY 532 -----IleGlnAlaAspAlaMetAlaAsnAlaTyrAlaIleGluLysGlyLys 547
Db 3187 TGGTGTGGTATGTTCTTAAGATAATGCTGTAGTAAATCTTTGAAGCTGAAGGGA 3246
OY 547 sArgLeuPhePheAlaGlyLeuArgGluMetTyrProGlyArgAlaLeuProSerAspAl 567
Db 3247 TCACAACTATCTGCCA-----AAACAGATGACTAACAAGAAAGCTTCGACGTGTTTGT 3300
OY 567 aaSnPheThrMetArgMetSerTyrGlySerIleLysGlyTyrGluProGlnAsp----- 585
Db 3301 TATATGATGGCAATGGGATGACTTTCATATCAACTAGTGGTATCAAGCAAAACACGCTT 3360
OY 586 -----GlyAlaTyrPheAsnTyrHisThrThrGlyLysGlyValLeu----- 599
Db 3361 TGTTCAGATGCCAAAGAAACTGCTATCTTGCATATATATGCCCATATGCTTATAG 3420
OY 600 -----GluLysGlnAspProLysSerAspGluPheAlaValGlnGly 613
Db 3421 CTTCACAGACGATAAATGGGATGCAATACTTTTATCAATATGGTTCAATGCTGCA 3480
OY 613 uAsnIleLeuAsp-----LeuPheArgThrLysAsnTyrGlyArgTyrAlaGluAsnGlyLys 632

```

```

Db 3481 ATCTTCTTGGAAAAACGCTGATGGACAGCAAGACTATTTGGTCAATCTAGAAATAGATA 3540
OY 632 nLeuHisIleAlaPheLeuSerAsnAspIleThrGlyLysAsnSerGlySerProVa 652
Db 3541 TAGTAAATGGTTTATTCATTTGATTAATGAT-----AGTAAGTGGCGTTA 3585
OY 652 lPheAspLysAsnGlyArgLeu---IleGlyLeu---AlaPheAspGlyAsnTrpGluAl 670
Db 3586 TTTTGATGCCAGTGGAGTCATGCTGTAGCTTTGAAACAAATTAACGCCAATACACAGTA 3645
OY 670 aMetSerGlyAsp 674
Db 3646 CTTCATGATCAAGAT 3658

RESULT 15
US-09-210-361-5
; Sequence 5, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; EARLIER FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 4310
; TYPE: DNA
; ORGANISM: streptococcus mutans
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (5)...(4295)
US-09-210-361-5

Alignment Scores:
Pred. No.: 0.736 Length: 4310
Score: 102.00 Matches: 154
Percent Similarity: 34.30% Conservative: 129
Best Local Similarity: 18.67% Mismatches: 280
Query Match: 2.74% Indels: 263
Gaps: 41

US-10-008-355-2 (1-712) x US-09-210-361-5 (1-4310)
OY 20 GlyValAlaLysAlaAspLysGlyMetThrPheLeuAsnGluLeuAsn----- 35
Db 1463 GGTGTGATTAAGATGAGAAAAT-----GCGATTATATCATCTTCCATTTAGAGCT 1516
OY 36 -----GlnGluAsnLeuAspArgMetArgGluLeuGlnGlyPheThrLeu 49
Db 1517 TGGTCAGATTAATGATCCCAATACAAATAAAGATTAAG-----GGTCACAAATTA 1567
OY 50 ProLeuAspSer-----LeuTyrSerPheAspLysPro----- 60
Db 1568 CCGATTGATTAATAAATCGCGCTATCGCTTATATGCTTTGACCGCTCTTGAAAAA 1627
OY 61 SerIleAlaAsnAlaValValIlePheGlyGlyGlyCysThrGlyIleThrValSerAsp 80
Db 1628 GATGCAAGCAATAAATGAATTCGACGCGGACTTGAACCTGTGATACAAATAGC--- 1684

```

QY 81 GlnGlyLeuIlePheThrAsnHisCysGlyTyrGlyAlaIleGlnSerGlnSerThr 100
Db 1685 -----TTGAATTAACGCTTCAGCTGAAGGTAA--AAATGTGAACGTATG 1726
QY 101 ValAspHisAspTyrIleuArg-----AspGlyPheValSerArgThrMetGlyGlu 118
Db 1727 GCACACATATTTTATATCCCGCTCAGACAGCAAGTCCAAACGGT----- 1774
QY 119 LeuProIleProGlyLeuSerValTyrTyrLeuArgLysIleValLysValThrAspLys 138
Db 1775 -----ATTGCTAAATTAATTAAAGCT-----CAG 1798
QY 139 ValGlnGlyGlnLeuLysGlyIleThrAspIleMetGlnArgLeuArgLysAlaGlnGlu 158
Db 1799 ATTAAATCCCAAAACAGATGGTTGACCTTACTTGGATGAATGAAGCAAGCCTTTAAG 1858
QY 159 ValCysGlnGlu-----LeuAlaLysLysGlnAsnAlaAspGlnAsnGlnLeuCys 175
Db 1859 ATCTACAAATGAAGCATGCGTCAGCTACAGAAAAGTACACACATCCCATATTCGCACA 1918
QY 176 IleValGlnProPheTyrSerAsnAsnGlnTyrPheLeuIleValTyr-----AspVal 193
Db 1919 GCCTATGCTTGGATGCTGCCAATTAAGATTCTATACAGCTCTTATATGTGATATG 1978
QY 194 PheLysAspValArgMetValPheAlaProPheSer----- 205
Db 1979 TACAGTATGACGCTCATATATATGGCGACTAAATCCCTTATATATGCTATTTGATCT 2038
QY 206 -----SerValGlyLysPhe 210
Db 2039 TTATTAAGCAGCTATTAATATATGCCGGTGTCAAGACATGAAGATCACTATGTT 2098
QY 211 GlyGlyAspHisAspAsnTyrPheTyrProArgHisThrGlyAspPheSerValPheArg 230
Db 2099 GAAGGTATATAAAGCTATATGATGGAT--TATACAGCGCTTTGACTTGTGTGCT 2155
QY 231 ValTyrAlaGlyAlaAsp----- 236
Db 2156 TATGTGTACAGAGCTATATGAAGCTACAGATCAAGCGAGTACACTAAACACAAAGA 2215
QY 237 -----AsnArgProAla----- 240
Db 2216 ATGCGCTCATTTACACACATTAACCTTAGCCTTAATGAATCAATAATGAATAGTAT 2275
QY 241 -----GlnTyrSerLysAspAsnLysProTyrLysProValTyrPheAlaIleVal 257
Db 2276 GTCAATATGGGGCTCGCATTAATAATCAAGAGTACCGCTCCGCTCTTAACAACATAA 2335
QY 258 ---SerMetGlnGlyTyrLysAlaAspAspTyrLamethrIleGlyPheProGlySer 276
Db 2336 GATGGTTTGACAAAGCTACACTCTGTGATGCCGCTGCTAAATCCCTT--TATCGCAAAAG 2392
QY 277 ThrAspArgTyrLeuThrIleSerTyrGlyValGlnAspArgIleGlnAsnGlnLysAsnPro 296
Db 2393 AATGATTAAGGAGAAATAGCTTTGATGCTAGTAC--ATTCAAGGTACTCTAAATCCG 2449
QY 297 ArgIleGlnValArgGlyIleLysGlnGlyIleTyrLysGlnAlaMetSerAlaAsp-- 315
Db 2450 -----CAAGTATCAGGTAT--TTAGCCGTTGGGTCCAGTAGAGAGTATGATAT 2500
QY 316 GlnAlaIleThrArgIleLysTyrAlaSerLysTyrLamethrIleAsnIleAsnTyrTyrLysAsn 335
Db 2501 CAAAGATTTGCTTAGACAGCAAGCAATTAAGCAAAATGCTATGTGTCAAGTCTACGAATCA 2550
QY 336 SerIleGlyMetAsn-----ArgGlyLeuAlaArgLeu----- 346
Db 2561 TCAAGTGTCTGTGATTTGCAATTTATAGCAAGGTTTCTCAAACTTCCAAAGATTTTGT 2620
QY 347 -----AspValIleGlyArgLysAlaIle----- 354
Db 2621 ACAGAAATTTACAGACTATTAATTAAGAGATTTCTCAAAATGTCAAACTTCTCAAAATCT 2680

QY 355 -----GlnGluArgAlaPhe 359
Db 2681 TGGGGTCTCACTTCCTTTGAATGGCACCCCAATATGCTCTTCTGAGATGGTTCTTTT 2740
QY 360 AlaAspTyrIleArgLysAsnGly----- 367
Db 2741 CTAGACCTATTTATCAAAATGGTATGCTTGAAGATGGTATATGATCTGTGATAGT 2800
QY 368 LysSerAlaValTyrGly-----AspValLeuSerSerLeuGlnLysAlaTyrLys 384
Db 2801 AAGATTAACAAATACGCTTCTCAGCAAGACATGATTTATGCAGTTAAAGCTCTGCATAAA 2860
QY 385 GlnGlyAlaLysAlaAsnArgLysMetThrTyrLeuSerGlnThrLeuPheGly----- 402
Db 2861 AGCGGTATTCAGGTTATTCGGGAT--TGGGTACCAAGTCAAACTTAATCTTCCG 2914
QY 403 GlyTyrGlnValValArgPheAlaGlnPheAlaAsnAlaLeuAlaThrAsnProAspAla 422
Db 2915 GGCAAAGAAGTCGTA-----ACGGCTACACGTGCACAGT 2950
QY 423 His-AlaGlyIleLeuLysSerLeuAspLysTyrLysAspTyrLeuProSerLeuAs 442
Db 2951 TATGTGATGATCGCAAAAGCTCGAATCAAAATACACTATATGCTGCCACA----- 3005
QY 442 PArgLysValLeuProAlaMetLeuAspIleValArgArgIleProAlaAspLysLe 462
Db 3006 -----CTAAGACTAATGTAGAGATTTATCAAGCGAAGTATG----- 3041
QY 462 uProAspIlePheLysAsnValIleAspLysLysPheLysGlyAspThrLysLysTyrAl 482
Db 3041 ----- 3041
QY 482 aAspPheValPheAspLysSerValValProTyrSerAspLysPheIleAlaMetLeuLys 502
Db 3042 -----CGCGTCTTCCTTAGTACGACCTCGCTGATACCTACT 3082
QY 502 sSerMetAspLysGlnLysPheAlaLysAlaIleGlnLysAspProAlaValGlnLeuSe 522
Db 3083 ATCTTTAACCGCAGCGCAAAATTTCAAAATGTATAG-AAAGTTGATCAAGCGCAAAA----- 3136
QY 522 rLysSerValIleAlaAlaAlaArgAla----- 531
Db 3137 -----ATCACACAGATGAAGCAAAATACTCATATGGGCAAAATATTTAGGCCG 3186
QY 532 -----IleGlnAlaAspAlaMetAlaAsnAlaTyrAlaIleGlnLysGlyLys 547
Db 3187 TGGTGTGTTATGCTTCTTAAGATATGCTAGATGAATACTTTGAACSTGAAGAGGAA 3246
QY 547 sArgLeuPhePheAlaGlyLeuArgLysMetTyrProGlyValArgLalaLeuProSerAspAl 567
Db 3247 TCAAACCTATCTGCCA-----AAACAGATGACTAACAAAGAACCTTCGACGTGTTTGT 3300
QY 567 aAsnPheThrMetArgMetSerTyrGlySerIleLysGlyTyrGlnProGlnAsp----- 585
Db 3301 TAAATGATGGCAATGAGTACTTCTATTCACATGAGGTATATCAACCAAGAACACTT 3360
QY 586 -----GlnAlaTyrTyrAsnTyrHisThrThrGlyLysGlyValLeu-- 599
Db 3361 TGTTCAGATGSCAAAGAAATGCTATCTTGTATGAATATGATGSCCATATGCTTATG 3420
QY 600 -----GlnLysGlnAspProLysSerAspGlnIleAlaValAlaGlnI 613
Db 3421 CTTAACAGACSTAAATGCGCAAGTGCATTTTATCAAAATGCTGTCAATTTGCGCTGA 3480
QY 613 uAsnIleLeuAsp---LeuPheArgThrLysAsnTyrGlyValTyrGlyAlaGlnGlnGly 632
Db 3481 ATCTTTCTTGGAAACCCCTGATGSCACAGCAAGATATTTTGTGATCTAGAAATATGATA 3540
QY 632 pLeuHisIleAlaPheLeuSerAsnAsnAspIleThrGlyGlyLysSerGlySerProva 652
Db 3541 TAGTAAGTGGTATGATTCATTTGATTAATGAT-----AGTAAGTGGCGCTTA 3585
QY 652 lPheAspLysAsnGlnArgLeu---IleGlyLeu---AlaPheAspGlnLysnTyrGlnAl 670

Db	3586	TTTTGATGCCAGTGGAGTCATGGCTGTAGCTTTGAAAACATTAAAGGCATATACGCAGTA	3645
Qy	670	ametSergIyasp	674
Db	3646	CFTTGATCAAGAT	3658

Search completed: October 18, 2002, 01:09:38
Job time : 4113 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 17, 2002, 22:28:31 ; Search time 43 Seconds
(without alignments)
1591.061 Million cell updates/sec

Title: US-10-008-355-2
Perfect score: 3719
Sequence: 1 MOKRLSILGALLGASG.....LFMIDKMGQCPRLIQELKLI 712

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	907.5	24.4	716	2	G82627 hypothetical prote
2	133.5	3.6	1104	2	alpha-amylase (EC
3	133	3.6	883	2	D97933 valine--tRNA ligas
4	129	3.5	1104	1	A36866 microbial collagen
5	125	3.4	1440	2	T33813 hypothetical prote
6	124.5	3.3	735	2	F98228 1,4-alpha-glucan b
7	124.5	3.3	735	2	AH3057 glycozen branching
8	123.5	3.3	776	2	S44784 C30C11.4 protein -
9	123	3.3	1289	2	F72308 hypothetical prote
10	122.5	3.3	1571	2	AC1647 Lactobacillus phag
11	122	3.3	721	2	AD1617 penicillin-binding
12	122	3.3	743	2	D82883 DNA helicase II uti
13	122	3.3	883	2	C95066 DNA helicase (EC 3
14	120	3.2	666	1	D65103 valyl-tRNA synthet
15	118	3.2	666	1	C82332 translation elonga
16	116	3.1	952	2	T36664 probable DNA topoi
17	115.5	3.1	1154	2	F71856 hypothetical prote
18	114.5	3.1	561	2	G83913 hypothetical prote
19	114.5	3.1	1198	2	T42223 ladder protein - C
20	114.5	3.1	1306	2	S22624 glutamate synthase
21	114.5	3.1	1530	2	H83865 hypothetical prote
22	114.5	3.1	1758	2	T24393 dnaK-type molecu
23	114	3.1	620	2	S19660 dnaK-type molecu
24	114	3.1	620	2	S73236 transferrin-bindin
25	114	3.1	711	2	S70660 internal virion pr
26	114	3.1	747	1	H1BPC7 DNA topoisomerase
27	113.5	3.0	759	2	A61686 pyruvate formate-1
28	113	3.0	1047	2	F81728 ribonucleoside-dip
29	113	3.0	1047	2	F81728

30	112.5	3.0	657	1	A64079 2',3'-cyclic-nucle
31	112.5	3.0	767	2	G81064 topoisomerase IV c
32	112.5	3.0	957	2	C69463 type I restriction
33	112	3.0	554	2	T31783 type I restriction
34	112	3.0	557	2	A56690 esterase - Caenoth
35	112	3.0	804	2	T49975 hypothetical prote
36	112	3.0	1504	2	A33602 DNA-directed DNA p
37	111.5	3.0	878	2	A71641 aconitate hydratase
38	111.5	3.0	922	2	T18878 hypothetical prote
39	111.5	3.0	1164	2	G71827 hypothetical prote
40	111	3.0	415	2	AE3522 Leu/ile/Val-bindin
41	111	3.0	1092	2	S42798 fibronectin-bindin
42	111	3.0	1167	2	AF0258 probable phage tai
43	110.5	3.0	992	2	S54396 protein L precursor
44	110.5	3.0	1039	2	C64418 isoleucine--tRNA 1
45	110	3.0	1339	2	H81307 restriction modifi

ALIGNMENTS

RESULT 1
G82627 hypothetical protein XF1887 [imported] - Xylella fastidiosa (strain 9a5c)
C:Species: Xylella fastidiosa
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: G82627
R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82515; MUID:20365717
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: G82627
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-716 <SIM>
A:Cross-references: GB:AE004008; GB:AE003849; NID:g9106961; PIDN:AAF4693.1; GSPDB:GN
A:Experimental source: strain 9a5c
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.
Britones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carr
as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fr
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; La
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmeti,
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sava
M.; Tenuhko, M.H.; Vallada, H.; Van Sluys, M.A.; Vertovski-Almeida, S.; Vettore, A.L.
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XF1887
Query Match 24.4%; Score 907.5; DB 2; Length 716;
Best Local Similarity 31.4%; Pred. No. 4.6e-51;
Matches 239; Conservative 135; Mismatches 291; Indels 95; Gaps 21;
1 MOKRLSILGALLGASGVAKADKGMILNENENLDRMRELGFYLPISLVSFFDKP 60
1 MKNRLSLSLVATLITVDS--THAGSGMVPPOOL-PEINGPLKQAGLOLSPEGLSLUTGD 57
61 STANAVVIFGGCGTGITVSODGLIFTNHCGAGATOSSTVDHYLRDGFVSFTMGEELP 120
58 PMGAAVVSLL--GNCTASLVSPEGLVITNHHCAVGAIDNSTPKNKLKEGFNALQADEVS 115
121 I-PGLSVKTLKRTIKVTKDVEGOLKGIPTDEMERTKRAQVCC--QELAKENNDEMOLCTV 177
116 AGNNARIVYLEDTDTVTAQAKALAAAGNDPFRRTALTSTFSKOETAKCEE--EQGRCOF 174
178 EPFYSNNEFYFLIYVDVFKVDMVFAPSSVGRFGGDTDMMMPRHGTGSVFRVYGAQN 237

```

Db 175 FSFAGNTYRVFKNLEIKDVLRYAPQSGVCKFGGDVNMMPRHGTDFSEYRAYVGKDG 234
OY 238 RPAEYKONKRYKPYFAAAMOGYKADDDYAMTIGFPGSTDRYLSMVEEDRIENENPR 297
Db 235 KPASCKENIYRPHKMLKESDQPLGDDFWAGYVAGERTNRVALY-----AEENIAH 288
OY 298 IEVNGIKO-----GIMKEASADQATRIKASYKSAOASANYKNSIGMNGRLARLDVIGR 351
Db 289 WTYPIVIGHFNLTALIEAASKQNDIOVKYASTLAGNNTSKNPDGQDQDFRRINAIQ 348
OY 352 KRAEBRPAADIRKNG-----KSAYIGDVLSSLEKAYKGAKANREMYLSELTLEGEYV 407
Db 349 KOSEETAVLWMLKQOQIGHEBALAHQTLVDLTEOY-----KANODRDFVLGO-FENGSGVI 403
OY 408 RPAQFANLA---TNPDAH-----AGILKSLDKYKDYLSLRKY-----445
Db 404 GVAANVLYRLAERTKSDQREAGYOERDLPTIEGMLKOME---RRYLDPMORQMOYWT 460
OY 446 ---LPAMLDIVRRIRPA-----DKLPDIFKNYIDKKFGKDTKKYADFVEDKSVYPSD 495
Db 461 EYNKLP-----VKQRYVALDVWLDGIPATLKLRLGDTKLSSSEER-----500
OY 496 KFHAMLSKMDKEKFAKAIKDPAPAVELSKSVIAAARAIQADMANAYAEKGRLEFAGLR 555
Db 501 ---LKFENADRAAFESSODPAIRYAVAIIMPALLEIERONKIRTGELKARPYTLQALA 555
OY 556 EMYP---GRALPSDANFTRMYSIGYEPDQGANYNHTTGKGLKODKSPDEFAQOE 613
Db 556 DYNSHGFEVYPDANSIRITFGYHAKGSPKDYETPTFLQGYMAK-NGVEFEDSPK 614
OY 614 NILDFTKNGRYAEN---GOLHIAFLSNNDITGNSGSPYEDKNGRLIGLAFDGNWEAM 671
Db 615 SLINAKKASKYANLADQIGTVPVAFSLSDLDITGNSGSPYEDAHGKLVGLAFDGNMESV 674
OY 672 SGDIIEFPDLORTISVDIRYVLEMDKGCQCPRLIOELK 711
Db 675 SSNWFDPMWTRTIAVDSRYVQWMTTEVAPAPHLKLELNL 714

```

RESULT 2

```

A60999
alpha-amylase (EC 3.2.1.1) precursor - Micrococcus sp. (strain 207)
C:Species: Micrococcus sp.
C>Date: 31-Dec-1993 #sequence.revision 31-Dec-1993 #text.change 15-Oct-1999
C:Accession: A60999
R:Kimura, T.; Horikoshi, K.
FEMS Microbiol. Lett. 71, 35-42, 1990
A:Title: The nucleotide sequence of an alpha-amylase gene from an alkalophilic
A:Reference number: A60999
A:Accession: A60999
A:Molecule type: DNA
A:Residues: 1-1104 <KIM>
A:Cross-references: GB:X55799; NID:g296762; PIDN:CAA39321.1; PID:g296763
C:Function:
A:Description: catalyzes the hydrolysis of internal 1,4-alpha-D-glucosidic bonds
A:Pathway: glycogen/starch degradation
C:Keywords: glycosidase; hydrolase; polysaccharide degradation
F:1.32/Domain: signal sequence #status predicted <SIG>
F:3.3-1104/Product: alpha-amylase #status predicted <MAT>

```

```

Query Match 3.6%; Score 133.5; DB 2; Length 1104;
Best Local Similarity 20.8%; Pred. No. 1.8;
Matches 138; Conservative 94; Mismatches 259; Indels 173; Gaps 35;

```

```

OY 141 GOLGIDTEMERLKAQVCEQLAKKENADENQICIVPEPFSNNEXFLIVDFEDVMY 200
Db 384 GDFGIGTIDKLTYL-----DELGVNTIMI-SPVENIKY-----DVRXY 420
OY 201 FAPSSVCKFGGDTDNM---MMPRHTGDSVFR-VYAGADNRPAXSKDKPYKPYFAAV 257
Db 421 ETSSEYIYRHGYWANNFGLNP-HGTMEEFHDLIDGADHRNMKIMVD-----VVV 470
OY 258 SMOGY-KADDYAMT---IGFPGSTDRYLS-----MGVEDRIENEN 294

```

```

Db 471 NHTGGLKEIDGSVTPNPAGYPSDADRARFSDLLRGADVGTDDEVGELACLPDFITEEDP 530
OY 295 NPRIEVRKIKGIMKEAMSADQATRIKY-----ASKYASANYKNSIGMNGRLARLDVIG 350
Db 531 NVRKQIID-WQTDWIEKATYENGNTIDFVRVDTQKHVEDAWMOFKNALTERKMEPEFKMG 589
OY 351 KRAEBRPAADIRKNGSAVY---GDVLSLEKAYKGAK-----ANREMYLSE 398
Db 590 EA-----WGAKVDNTGLTLETGYMDSLIDGCFKETAASFNGSLEAANASLTARNA 640
OY 399 TLFGTEYVRP-----AQFANALATN-----PDHAGILKSLDKYK 435
Db 641 KLDWTATLGGFLGSHDEBGLFSLAGDKQLQVATLQATKAGQPVLYYG---EELGOTGA 698
OY 436 DYLSLDRKVLPAHLDIVRRIRPADKLPDITKNTIDKKFKDCTKKYADFVEDKSVYPSD 495
Db 699 NNYPQYDNR---YDPAWDVEGENEILAHYTKILNR-EQYSKYFAK--GERTLVGSD 750
OY 496 KFHAMLSKMDKEKFAKAIKDPAPAVELSKSV---IAAARAIQADMANA-YAIEKGRLEF 551
Db 751 KDQFLTSRDYQDQKVVYGLVAEE-SKAVTLTVDSADAVTDAVSGEYRATAGK-----805
OY 552 AGLREMPGRALPSDANFTRMYSIGYK-----YEPDQGANYN 590
Db 806 ---VNLTLPGKADPGTVLLT---VEGGINITGVAKDGEVYVELVPENNIRIRHKREDNMYKN 861
OY 591 YHTGKGLKLEKODKSPDEFAVOENILDFRTKNGRYAE-----NGOLHIAFLSNNDITGG 646
Db 862 Y---GAMLMNDVSPSANMPGATMFE--KTDSTGATIDVLDAGKAKNIGLV-MDTITAG 915
OY 647 NSGSPVEDKNGRLIGLAFDGNWEAMSGD--TEFEP-DLORTISVDIRYVLEMD-----697
Db 916 DAGKDGDKGFTISSPQANEIWIKQSGDKVYTYEPVLDLPAN-TYRIHYTRBAVYDDPGI 974
OY 698 -KKG 700
Db 975 WNWG 978

```

RESULT 3

```

D97933
valine--tRNA ligase (EC 6.1.1.9) [imported] - Streptococcus pneumoniae (strain R6)
C:Species: Streptococcus pneumoniae
C>Date: 22-Oct-2001 #sequence.revision 22-Oct-2001 #text.change 02-Nov-2001
C:Accession: D97933
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; Dehoff, B.S.;
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.;
Y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: D97933
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-883 <KUR>
A:Cross-references: GB:AE07317; PIDN:AAK9296.1; PID:g15458063; GSPDB:GN00174
C:Genetics:
A:Gene: valS
C:Superfamily: valine--tRNA ligase
C:Keywords: ligase

```

```

Query Match 3.6%; Score 133; DB 2; Length 883;
Best Local Similarity 20.4%; Pred. No. 1.4;
Matches 104; Conservative 65; Mismatches 166; Indels 176; Gaps 24;

```

```

OY 192 DVF-----KDYRMVFAPPSSVCKFGGDTDNMMPRHGTGDSVFRVYAGADNRPAEYS 243
Db 26 DVFRPSGQKAKKPYSIYIPPNVYTKL-----HLG-----HA 57
OY 244 KDNKPYKPYFAAVSMQYKADYAMTIGFPGSTDRYL-TSWGVEDRIENENNPRIEYRG 302

```

Db 58 WDT-TLDDIIIRKMGFD-----TLWLPBGMDHAGIATQAKVEERLRGEGISRYDLGR 110
 QY 303 -----IKGIMKEAMSAOATRIKYA-SKYASANYWKNISGMNGLARLDVIGKRAEER 357
 Db 111 ESFTLKWEMKD-----EYATTIKEQMGKGLSDVYSREPTFLDGLS-----KAVRK 158
 QY 358 AFADMIK-----NGKSAVYGD-----VLSSLEKAYK-----EGAKANEMT 394
 Db 159 VFVDLYKKGMYRGEFTIINMDPARFALSDIEVYHKDEGAFYMMNYMLEDSGSALEVAT 218
 QY 395 YLSTLEFGTEEVVRFQAFANALATNPDAHGIKSLDDKKDYLP-----LDRKVLPAW 449
 Db 219 TRBETMGDV-----AIAVNP-----DPRYKDLGKVVILPIANKLIPIY 259
 QY 450 LD-----IYRRRIIPADKLPDIFKNVIDKFKKDDTKKYPADVPVDPKSVYPSDK 496
 Db 260 GDEHADPEFTGYVKTIPADHPNDFLQGRHNPQVAINMDGDIMNELVEFSGM-----DR 316
 QY 497 FHMLKSMDEKEFAKIEKPAVELSKSVIAAARA-----IQAD 535
 Db 317 FEAR-----KAVAKLEIGALVYIEKRVHVSIGSERGVVPERLSTOMFVKMDOLAKN 371
 QY 536 AMANAVYIEKGRLEFAGLEMTPGRALPS-----DANFTMRMSYSGIKYEPQDG 586
 Db 372 AIANQDTEK-----VEFYPRFNDPFLQMMENVHDMVISROLWGH-----QIP 416
 QY 587 AMYVHTTGKGVLEKQDPKSDPEFAVOENIID 617
 Db 417 AMYN-ADGEMVYGEAPESGDGTWQDDEVID 445

RESULT 4
 A36866
 microbial collagenase (EC 3.4.24.3) precursor - Clostridium perfringens
 C:Species: Clostridium perfringens
 C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
 C/Accession: A36866
 R:Matsumita, O.; Yoshihara, K.; Katayama, S.I.; Minami, J.; Okabe, A.
 J. Bacteriol. 176, 149-156, 1994
 A:Title: Purification and characterization of a Clostridium perfringens 120-kilodalton
 A:Reference number: A36866; MVID:94110220
 A:Accession: A36866
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1104 <MAT>
 A:Cross-references: GB:D13791; NID:9440850; PIDN:BA02941.1; PID:9440851
 C:Genetics:
 A:Gene: COLA
 C:Superfamily: microbial collagenase
 C:Keywords: hydrolase; metalloproteinase

Query Match 3.58; Score 129; DB 1; Length 1104;
 Best Local Similarity 18.2%; Pred. No. 3.6;
 Matches 142; Conservative 106; Mismatches 276; Indels 258; Gaps 36;
 QY 29 WLNLNLEQENL-----DRMRELFTPLDLSYSPDKPSIANAVYFEGGCT---GIYVSD 80
 Db 419 WAKSEVAAQEMRVYQNDKALEGNPDILTVIYNSPEETKLNRIINGFSTDNCGIYEN 478
 QY 81 QGLFTNHHCYGAIGS-QSTVDH-----YLKDFVSRMGELPIPL---SVYELRKIV 133
 Db 479 IGFFFTYERTPEESIVYLEELFRHEFTHYLGGRV-----VPGMMGGEGYQGSVL 529
 QY 134 KYVDKVGQLKGTIDEMERLKAQVCOELAKKENADENQLCIYEPYSNNEYFLIYDV 193
 Db 530 TWVEEGTAEEFAGSTRDGIKPRKSVYQGLAYDRNNRMSLYGLHARYGSDPF-----583
 QY 194 FKDVRAFAPSSVYKGGGTDMNMPRHGTGDSVFERYVAGADRPAEYSDKNPKPYV 253
 Db 584 -----NKGFLSNMYNNKGMF-----NKMNYIKNN-----611
 QY 254 PAAVSMQGYK-----ADYAMTIGFPGSTDRYLTLSGVEDRIENENNPRIE---VGIGK 304

Db 612 ----DVSGYKDYIASMSDGYLNDKYODYMDSL-----NNIDNLDVPLVSEYVNG-- 659
 QY 305 QGIMKEAMSAOATRIKYA-SKYASANYWKNISGMNGLARLDVIGKRAEER 357
 Db 660 ----HEAKDINEITNDIKEVSNIKDLSNVEKSGFTTYDGRGYVGGRSQGEEN---DW 712
 QY 363 IRKNGKAVYGDVLSLEK-----AYKEGAKANREMYTSETLFGTEEVVRFQAF 412
 Db 713 KDMASK---LNDMLKELSKSMNGYKVTAFVHHKXDEKNGNYDDVYFHGMNT-----763
 QY 413 ANALATNPDAHA-----GIKSLDDKYKDYLPISDRKVLPAMLDIYRRRIIPADKLPDIFK 467
 Db 764 ----DTWTDVHVNKEPKAVIKS-----DSSV-----IYEEIIN-----FD 794
 QY 468 NVIDKFKRGDTKKY-----ADFEVPEKSVYPSDKFHALKSM-----K 506
 Db 795 GTESKDDDEGELKAYEMDFGDEKSNKAKHKKYK-----TGEYEVKLYTDNNGGINT 849
 QY 507 EKFAKALEKDP-----AVELSKSVIAAARAIOADAMANAAYIEKGR--- 548
 Db 850 SKKIKVYEDKPEVYINSEPNDEKANOJAKSMMLYKGTISEDYSDKYTFVAKKGNV 909
 QY 549 -----LFFAGLREMY-----PGRALPSDANFTMRMSYSGIKYEPQDG 586
 Db 910 KITLNLNSVGIPTWTLTKEGDLNMYLYANGNDGTELKGETLPGRYLYLSVYTDNQS 969
 QY 587 AMYVHTTGKGVLEKQDPKSDPEFAVOENIIDLFTKRYGRAENGQHLIAFLSNNDITG 646
 Db 970 A--YTVANVGMLNKEVETEKEKAKE-----VEN-----NND-----999
 QY 647 NSGSPVFPK-----NGRLIG-LAFDGNWEMASGDIPEPDLQRTI---SVDIRYVLM 695
 Db 1000 -----FDKAMKVDNSKIVGTTLSNDLKDLYSIDIONPSDLNIVENLDNIKMMILYS 1053
 QY 696 ID 697
 Db 1054 AD 1055

RESULT 5
 T33813
 hypothetical protein VC5.3 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
 C/Accession: T33813
 R:Tin-Wollam, A.M.; Woldmann, P.
 submitted to the EMBL Data Library, November 1998
 A:Description: The sequence of C. elegans cosmid VC5.
 A:Reference number: 221417
 A:Accession: T33813
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1440 <TIN>
 A:Cross-references: EMBL:AF106581; PIDN:AACT8208.1; GSPDB:GN00023; CESP:VC5.3
 A:Experimental source: strain Bristol N2; clone VC5
 C:Genetics:
 A:Gene: CESP:VC5.3
 A:Map position: 5
 A:Introns: 15/3; 27/2; 258/3; 382/3; 506/3; 630/3; 754/3; 879/3; 1005/3; 1410/3
 Query Match 3.48; Score 125; DB 2; Length 1440;
 Best Local Similarity 20.1%; Pred. No. 9.8;
 Matches 108; Conservative 72; Mismatches 162; Indels 194; Gaps 30;
 QY 132 IVK-VTDKVEGQLKGTIDEMERLKAQVCOELAKKENADENQLCIYEPYSNNEYFLIV 190
 Db 237 LVKGVVDRODGEVYVIAEKM-----LSVCGEYVKESTRKR-----271
 QY 191 YDVKDVRAMFAPRSSVYKGGGTDMNMPRHGTGDSVFERYVAGADRPAEYSDKNPKY 250
 Db 272 ----RELEAFO-----DFVKKWTPQGLDITDALA-AKGDSEVQAKVE-----311
 QY 251 PVYPAVSMQGYKADYAMTIGFPGSTDRYLTLSGVEDRIENENNPRIEYVGIKGIWKE 310

[illegible]

RESULT 6
F98228
1,4-alpha-glucan branching enzyme (glycogen branching enzyme) [imported] - Agrobacterium
C:Species: Agrobacterium tumefaciens
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 11-Jan-2002

R:Goodner, B., Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quirillo, B.; Goldman, A.; Liu, F.; Mollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Merkleiz, B.; Science 294, 2223-2328, 2001

A:Title: Genome sequence of the Plant Pathogen and Biotechnology Agent *Agrobacterium tumefaciens*

A:Reference number: A97359; PMID:11743194

A:Accession: F98228

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-735 <KUR>

A:Cross-references: GB:AE007870; PIDD:NAK89352.1; PIDD:g15159198; GSFDB:GN00170

C:Genetics:

A:Gene: AGR_L1558

A:Map position: linear chromosome

C:Superfamily: 1,4-alpha-glucan branching enzyme

Query Match	3.3%	Score 124.5	DB 2	Length 735
Best Local Similarity	20.4%	Pred. NO. 3.9		
Matches 157; Conservative	83;	Mismatches	255;	Indels 275; Gaps 44;

```

Oy 108 DGFVSRFM---GEELPIGLSLVKYIKVIKYATDKVEGOLKGTITMEMLRKAEVCEOLA 164
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 43 EESASCFIPGAEEVSYLTLDOGNFVELKQIDP--DGFEEGIDLSKQOPVRYRACR--- 97
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 165 KKNENDENOLCIVEEFYEN-----NEYF-----LIYDYVEFKDVRM----- 199
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 98 ----DDAEMAVTIDPISFGVPLGPMDDYFVREGSHLRFLFD-----RMGAPHLKLEGEVF 147
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 200 ---VFAPSSVGKREGCDIDNNMMPRH-----TDESEVFR--YVAGADNRPAEYSKDNK 247
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 148 HFAVAPAPARARVSVYGDENNNMDCGRHRHVMFRKDTGIWEIFADVYAAG----- 195
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 248 PYKPYVFAAVSMOG---YKADDYA-----WTIEFPGSTDRYLTSMGVEDRIENEN 294
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 196 AVK---FELGANGSLLEPKADPYARGBELRPNKNSVYAPELTQW-----EDGAHREN 246
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 295 NPRIEVR-----GIKQGIWKEA-----MSADQ--ATRIKYASK----- 325
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 247 MAQVODRQAPISIVEYVANGSMQRREDGTFLSWDELAAOLIPICDIMEGFTTHIEFLITEHP 306

```

```

OY 326 -----YQASAWYKNSIGMNRGLAFLDYGKRAEBAFAFDMI----- 363
Db 307 YDPSMGYOTGLIYATARF-----GDPEGARF-VNGAHKVGIGVLLDVAHAFPTDEHG 360
OY 364 -----RNGKSAVYG-----DVLSL-----EKAYKEGAK-AN 390
Db 361 LRMFDGTALEYHADPROGFHPDMWTALYNFGRIEWMSYLNNALYMAEKKPHLOGLRADAV 420
OY 391 REMTYLSETL-----FGG--TEYVRRARQAFANALATIPDAHAGIAGLSDOKYKDYL 438
Db 421 ASMLDLDSRKEGEWIPNEYGRENLESVRLQKNNSLYVG--THPGWMTAESTS--W 476
OY 439 PSLDERKVPAMLDIVRRIRPADKLPIDIKANYIDKKFKGDDTKKAYADFYDVKSVYVSDKF 498
Db 477 PKVSPVHNEGGLG-----FGFKMNGFMHDLISY----FSRE-FVHKRFH 516
OY 499 AM-----LKSMDKEKFAKAIERKPAVELSKSVIAAARAIAQADAMANAVAL-----EKKG 547
Db 517 HOELTFGLLYAFTEFNVLPLSHDEVVHGGSLIKMSGDDMQKFAANLSYGYFMNGYGF 576
OY 548 RLFFAG-----LRE--MYPG-RALPSDANFMRMSYG-SIKGYEPD 585
Db 577 KILFNGGEFAOMSEWSEKSLDMNLRQYPMHEGMRRLVRDLNLYRSAAALHARDCEPDG 636
OY 586 GAM--YNYHTTGKYLEKQDKPSDEFAVOENILDLFERTKNY-----GRYAENGOLHIAF 637
Db 637 FFWLVVDHENSVPFAMLTARTAGEKPAVICNLTPYRENYVPLPVAGRMRE----- 688
OY 638 LSNND--TTGNGSSGPFVDKNGRLGLAFDQNGWEM-----SGDIEFPD 680
Db 669 ILNTDAEIIYGG-SGK--GNGGRVQAVDAGEIGAMLVPLPLATITMEPE 734

```

RESULT 7

glycogen branching enzyme g1gb [imported] - Agrobacterium tumefaciens (strain C58, Du
C:Species: Agrobacterium tumefaciens
C:date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 01-Feb-2002
C:accession: AH3057
R:wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo
erage, G.; Gillet, P.; Grant, C.; Guentherer, D.; Kutlyavin, T.; Levy, R.; Li, M.; McEl
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam
ster, E.W.
A:title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:reference_number: AB2577; PMID:11743193
A:accession: AH3057
A:status: preliminary
A:molecule_type: DNA
A:residues: 1-735 <KUR>
A:cross-references: GB:AE008689; PIDN:AL44878..1; PID:g17742527; GSPDB:GN00187
A:experimental_source: strain C58 (Dupont)
C:genetics:
A:gene: g1gb
A:map_position: linear chromosome
C:superfamily: 1,4-alpha-glucan branching enzyme

Query Match	3.3%	Score 124.5	DB 2	Length 735
Best Local Similarly	20.4%	Pred. No. 3.9		
Matches 157; Conservative	83;	Mismatches 255;	Indels 275;	Gaps 44;

OY	DGFVSRIM---GEEIPIGLSVKTIUKTKVYTKDKVEGOLKQITTBEMFLRKAQEOCELA	164
	: : : : : : : : : : : : : :	
Db	EGRSARCFIPAAEEFVSVLTLDGNFEVGELKQIDP--DGFFEGRIDLSRKQPVRACR---	97
	: : : : : : : : : : : : : :	
OY	KKENADEMQLIVEEFYSN-----NEFY-----LIIVDFKDVRM-----	199
	: : : : : : : : : : : : : :	
Db	-----DDAEAVATDPYSGCPVLGPMDDYFAREGSHLRITFD-----RMGAHPLKLEGEGE	147
	: : : : : : : : : : : : : :	
OY	200 ---VFAPSSYGKKGGGDNDMMMPRH-----TGDSVER--YYAGADNPALRESDNK	247
	: : : : : : : : : : : : : :	
Db	HFAVAWPAARARYVVDGDNMMNDGRHRYAREFRKDGIWIETIFAPDVAYAG-----	195
	: : : : : : : : : : : : : :	

QY 248 PYKPYFAAASMOG-----YKADYA-----MTIGFPGSTRYLTSWGVEDRIENEN 294
 Db 196 AYK---FELLGANGELLPLRADPARGRELPRKNAVSAPBLTKW-----EDQAREH 246
 QY 295 NPRIEVR-----GIKQIKWEA-----MSAQ--ATRIKAYSK----- 325
 Db 247 WAQYDQGRQPISTIYEIVAGSNQREDGTFLSWDELAQOLPYCDMGFTHEFLPIEHP 306
 QY 326 -----YASANYWKNSIGMNRGLARLDVIGRRKRAEERAFADWI----- 363
 Db 307 YDPSMGYQTLGLYAPARF-----GDEGPARF--VNGAHKKGIGVLLDWVPAHPTDEHG 360
 QY 364 -----RKNKSAYYG-----DVLSSL-----EKAKBEAK--AN 390
 Db 361 LRWFDGALIEHADPROGFHPDMNTATITNFGRIEVMSTLNNALYMAKEFHLDGLRVDV 420
 QY 391 REMTYLSETL-----FGG--TEVREAFANALATNPDAHGLSLDKYKDYL 438
 Db 421 ASMLYLDYSRKEGEMINENYGGRENLESVRLQKMSILYVG--THPGVMTAEESTS--W 476
 QY 439 PSLDRKVLPAMLDIVRRRIPADKLPIFKNVIDKKFKGDTKKYADFVFDKSVVPYSDKH 498
 Db 477 PKYSQPVHEGIG-----FGFKMNGFMHDTLSY---FSRE--PVHRRFH 516
 QY 499 AM-----LKSMDKKEKFAKATEKDPAVELSKSVIAAARAIGADANANAYAI-----EKGK 547
 Db 517 HOELTFGLLAFTENFVPLSHDEVHVGKSLAKMSGDDMKFANRISYGFMGYPGK 576
 QY 548 RLFEAG-----LRE--MYPG--RALPSDANFTMRMSYG--SINGYEPD 585
 Db 577 KLFMGGEFPAQWSEMSKSLDMNLROYPMHEGMRRLVROLNLTYSKAAHAMDCEPDG 636
 QY 586 GAW--YVYHTTGKVKLEKQDPKSDERAVQENIDLEFTKNT-----GRAENGOLHIAF 637
 Db 637 FRMLVVDHENSIVFAMRLTARGEKPVAVICNLTPYRENYVPLPVAGRMRE----- 688
 QY 638 LSND--ITGNGSGPVFDKNGKRLIGLAFDGNMEAM-----SGDIEEPD 680
 Db 689 ILNTDAELTGG-SGK---GNGGRVQAVDAGEIGAMLVPLPLATIMEPE 734

RESULT 8

C30C11.4 protein - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 29-Sep-1999
 C:Accession: S44784
 R:Favell, A.D.
 submitted to the EMBL Data Library, January 1993
 A:Description: Sequence of the C. elegans cosmid C30C11.
 A:Reference number: S44782
 A:Accession: S44784
 A>Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-776 <FAV>
 A:Cross-references: EMBL:J09634; NID:g156220; PIDN:AAA27967.1; PID:g156223
 C:Genetics: 36/2; 145/3; 255/1; 417/2; 732/3
 A:Introns: 36/2; 145/3; 255/1; 417/2; 732/3
 C:Superfamily: heat shock protein 91

Query Match 3.3%; Score 123.5; DB 2; Length 776;
 Best Local Similarity 21.1%; Pred. No. 4.9; Indels 145; Gaps 25;
 Matches 101; Conservative 82; Mismatches 151;

QY 118 ELPIPGLSVKYLRKIVYTDVVEGQ-----LKGIT-----DEM 150
 Db 283 QTPIP-LINIECFMEDKDYTGKMQROFEDLAAPFNRIKQVLLPMDGVSIXKEEIDEI 341
 QY 151 E-----RLKRAQVCOELAKE-----MADE-----NOLCIVEPPYSNNEFLIYD 192
 Db 342 EIVGSGSRIPMIREIVDLDFEGEKPTTMMODEAVARGAAMQCALISPTFVREF----- 395

QY 193 VFKD-----VRMVFAPPSSVKGEGDDNDMMRRHTGDS--VPRVYAGADNRPAREYSKD 245
 Db 396 AIKDQPIRIRLSV---NSTGENGENDVFS--PDEVFPFSLVLSLSPGPNVAHYAQP 451
 QY 246 N-KPKKPYFAAASMOGYKADDMYATIGFPGSTRYLTSWGVEDRIENENPRIEVRGK 304
 Db 452 NVYPHNVGHIGSMKVNAR-----PGAD-----GQNKV-----KKVAVNP 488
 QY 305 OGIKKEMASDAQATRIYAKSYAOSANYWKNSIGMNRGLARLDVIGRRKRAEERAF-ADWI 363
 Db 489 DGIPTIA-----SATMYEPRIVEEYPAEAMEVDGDAKTEAPAEPLPEV 531
 QY 364 RKNKSAYYGDVLSLEKAY-----KEGKANREMTYLSSETLFGTGEV 406
 Db 532 KTKLVPEVDEIVTISTIVSYDVQKFHMLEIQMOESDAREKAKADAKNS--LEEYIEYERDK 590
 QY 407 V--RFAOFANALATNPDAHGLILSLDKYKDYLPISDRKVLPAMLDIVRRRIPADKLPI 464
 Db 591 VSDQYAEFIIPAA--DEFPSVLSTEDDWLYDEGEDAER-----DYEKRL--SELKA 639
 QY 465 IFKNVIDKKKEGDTKKYADFVDSVY---PYS-----KFHAMLKSMDKEKFAKAE 514
 Db 640 VGPVVEREYRESEYRKPAFDSFDSIMRVKAYEDVANGGPTYAHLDSKEMEKVIYNAIE 698

RESULT 9

F72308
 hypothetical protein - Thermotoga maritima (strain MSB8)
 C:Species: Thermotoga maritima
 C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
 C:Accession: F72308
 C:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwin, M.L.; Dodson, R.J.; Haft, D.H.; Hic
 Garret, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,
 C.M.
 Nature 399, 323-329, 1999
 A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome
 A:Reference number: A72200; MUID:99287316
 A:Accession: F72308
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1289 <ARN>
 A:Cross-references: GB:AE001761; GB:AE00512; NID:g4981529; PIDN:AAD36071.1; PID:g498
 C:Genetics:
 A:Experimental source: strain MSB8
 A:Gene: TM0992

Query Match 3.3%; Score 123; DB 2; Length 1289;
 Best Local Similarity 18.1%; Pred. No. 11;
 Matches 125; Conservative 111; Mismatches 256; Indels 200; Gaps 29;

QY 102 DHDYLRGQFVSRTMGELPIFGLSVKYLRKIVKYTDVVEGQKQITP----- 148
 Db 22 DSDALR--YRLNTISE-----LNKSYLKAIONIVG--YNGLFKSGSDFFLETLYSNDPL 72
 QY 149 -----EMERLRKAOEVOCELAKKENADENOLCIVEPPYSNNEFL----- 188
 Db 73 ARYKESQOEARIKVQEARREGASNEELIKQLOLAQEARLRNKKOFIALIESNLNSIDYL 132
 QY 189 -----IYVDVFKDVAVPAPSSVGKFGGDDNDMMPRHTGDSV-FRYVA 233
 Db 133 ASETLKNLSLTKANKIIDVDVKNMKEILPLLE--GDFGVYIKNALIQLIDSYRKITFDYC 191
 QY 234 ----GADNRPAREYSKDKPKPYVFAAVSMOGYADYAMTIGPGSTDRLT--SGGYE 287
 Db 192 KANYGATEKIAQY-----WVKTYF-----VEPFSSSEKKVL-----DEVLTAADELK 235
 QY 288 DRINENNNPRIEVRGKOG-----IMKEA--MSADQATR-----IKYASKYAQSA 330
 Db 236 DRLKDRLETKLELIVSKGNALAKVIEETAKKASBSILKIIITPSLVELFRTKYINNV 295
 QY 331 NY--WKNSIGMNRGL--ARLDVIGRRKAE-----BRAFPADWIRKNGKSAVYGDVLSL 379
 Db 296 DFOITFNDIASNEVVFIKQIRELVGNDENEINRCYFPRAVYLSKKAAKKAINDVGEKT 355


```

      302 VSKYEGSKGKDLVLSVDFEFOKAVEDILKRNK-----GQKQVAGSD 343
      485 VEDKSVY-----PYSDKFHML--KSMDEKFAKAIKDPAVELSKVIAARAIOAD--- 535
      344 LDRRAEVVMDPYSGEVLALAGOKLNDKGEF-----DYSLGTFETAYAMGSVAKSSTLL 398
      536 -----AMAN-----AYAIKGRLLF-----AGLEMYGRRLPBDAN---FTMR 573
      399 GGMGCAIENKTVFTDPIALKGTKPKSSWFNRTGAGNPLDVGALETISSNYSMTQVAM 458
      574 SVGSIKGYEPQ-----DGAMYNHTTGKGYL-----EKODPKSDEFAVOENLL 616
      459 KMGGA-KYVNGPLRAPLSTPDDMRYYNOFGVKTGIDLEPEQGYKGGDOOTICK-IL 516
      617 DLFRTNRYGRYAENGOLHIAFLSNDDITGNSGSPVFDK-----NGRLIGLAFDGNME 669
      517 DF---AIGQVSYTPLOMAQVYSTIANGSGSRAPSMVEKIRNPSTNGDSTLATAN-- 570
      670 AMSGDIEPFDLORTISV---DIRYV 692
      571 -----EPKVLNKIGVSDNDIKTY 588

```

RESULT 12

```

D82883
DNA helicase II U0501 [imported] - Ureaplasma urealyticum
C:Species: Ureaplasma urealyticum
C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
C:Accession: D82883
R:Glass, J.I.; Lefkowitz, E.J.; Glass, J.S.; Helner, C.R.; Chen, E.Y.; Cassell, G.H.
submitted to Genbank, February 2000
A:Description: The complete sequence of Ureaplasma urealyticum. Alternate views of a mit
A:Reference number: AB2870
A:Accession: D82883
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-743 <GLA>
A:Cross-references: GB:AE002148; GB:AF222894; NID:g6699495; PIDN:AAF0913.1; GSPDB:GN001
A:Experimental source: serovar 3, biovar 1
A:Gene: uvrd; U0501
A:Genetic code: SGC3
C:Superfamily: helicase II

```

```

Query Match          3.3%; Score 122; DB 2; Length 743;
Best Local Similarity 17.7%; Pred. No. 5.7;
Matches 143; Conservative 129; Mismatches 281; Indels 256; Gaps 40;

```

```

      10 LGAAALLGASGVAKAD-----KGMWLNELNENLDNRRELGFLL---PLDSLXSPDKPSI 62
      22 LGQQLVIAAGTGCTGYVLLRLAYLITE--KNHPSRLIGFTFTKKADEM---KERY 74
      63 ANAVYI-----FGGCGTGTIVSDOGLIFTNHHCYGAIQSOATVDHDIYLRDGFVSRTM 115
      75 GKTIGVSTPYLSTFHSMC--VKILQDDIHLNHY---NNIKIIDTDQDEL-----LKEI 124
      116 GEELPLPGISVYKLRIVYVTKVEGOLKGTID-----EMERLKAQEV----- 159
      125 FQDLNTEKKS-QVIKKIKITISKVKNKFPQNDMLNEKNHKYLELVLDNDAQRLVDIYKI 183
      160 -COELAKKENADENOLC-----IVEFYSNNEFLVYVFKDVR----- 198
      184 YDRCERKLVYVDDDLINLTHKLFIEPFEVLEKQNKEDTILV--DEFQTNKIQYDLIS 241
      199 -----MFPAPPSVSGKEGSDTDMNMMPRHTGDSFVRYAGADNRPAYSKDN-K 247
      242 LATKHNQNLVVGDDPDQMIYSFSG-AEQWIIINFSQNK-----NTKYIILITNTR 291
      248 PYKPYTAAVSMQGYKADYAMTIGFPGSTDRYLTSMGVEDRIENENPRIEVRG---IK 304
      292 STQPIINTANRLIDANNNNY-----KKMLTAFNT-----NDNNLPITYLRGQNPID 336

```

```

      305 OGIM-----KEAMSDAOTRIKYSKAYOASANYMKNISGMNGLARLDVIGKRAE 355
      337 EAMWIAKRTRELLEEGTTPANQIAVLFRSNHYSRTIQ-----SMARESPYIILGSKRY 391
      356 ERA-FADWIRKNGKSAVYDVLSTLEKAYEGAKANREMYL-----SETLFGCEVVR 409
      392 ERAEIKDMI-----AYLKVNDLD-----ELSPRIINTPRRAIGPTFEHV 433
      410 AQA-----NALATNDAA-----AGILKSDDKKDYLPISLRKVLPA 448
      434 KHAYIINNLFEPALAEVERKNHILNNTQKNILNFVNLIKREIRDEMDL-----KIME 486
      449 MLDIVARRI-----PADKLPIFKNVIDKKEFGDTKRYAFVFPDKSVVPSDKFHAML 501
      487 ILELYKKAYENAYLLENKAEKDIDNVE--LKRMMKYVDRHDPDTINDYLNSTALYL 544
      502 KSMDEK-----FAKAIEKDP-AVELSKVIAARAIOADAMANAAYIEKRL 549
      545 NKDGKSKENVLLMTVHNSKGLEYNVFAGMNEGILLPSDRAINDDPK---GVEEERRI 601
      550 FFRAGL-----REMYPRALPBDANFTRMVSGSTIKGYEPQDGMRY 589
      602 AYVALTRAKKNLYISSACCYDPLARQVESPRIENIGENMKLIINSFKNKPED--- 657
      590 NYHTTGKYLEKODPKS--DEFAVOENIDLFTPKNYGRYAENGOLHIAFLSNDDITGCGN 647
      658 ---MPLKSLFKQDEERSWPDSDKOKKEVEDNFGQ-----TKND----- 693
      648 GSGSPVFDKNRGLIGLAF-DGNMEAMSGDI 675
      694 -----FEIGERYIVHTSFGDGVIIIGDGI 717

```

RESULT 13

```

C95066
valyl-tRNA synthetase [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C>Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001
C:Accession: C95066
R:Rietveld, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; H
on, J.D.; Mayhew, E.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radu, D.; Holtzapf
son, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venier, J.C.; Dougherty, B.A.; Morris
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; M01D:2157209; PMID:11463916.
A:Accession: C95066
A:Status: preliminary
A:Molecule type: DNA
A:Cross-references: GB:AE005672; PIDN:AAK74724.1; PID:q14972043; GSPDB:GN0164; TIGR:
A:Experimental source: strain TIGR4
A:Gene: SP0568
C:Superfamily: valine--tRNA ligase

```

```

Query Match          3.3%; Score 122; DB 2; Length 887;
Best Local Similarity 20.0%; Pred. No. 7.4;
Matches 102; Conservative 66; Mismatches 167; Indels 176; Gaps 24;

```

```

      192 DVF-----KQVRNVFAPPSVSGKEGSDTDMNMMPRHTGDSFVRYAGADNRPAYSK 243
      26 DVFPSGDOAKAKRYSTIVPPNVTGKL-----HLG-----HA 57
      244 KDNKPYKPYFAAVSMQGYKADYAMTIGFPGSTDRYL-TSMGVEDRIENENPRIEVRG 302
      58 WDT-TLQDIIIRQRKMGPD-----TLMLPGMDHAGIATQAKVEERLGEIGTFRDCLR 110
      303 ----IKQGIWKEAMSDAOTRIKYA-SKYAQSANTYKNSIGMNGIARLDVIGKRAEER 357
      111 ESFLTFVWEMKQD---EVAITIKQWGMKGLSDYGRERETFDIGIS-----KAVKR 158
      358 AFAWDWIRK-----NGKSAVYGD-----VLSSLEKAYK-----EGAKANREMT 394

```

Db 159 VFVNLYKGGWYIRGEFIINMPDPAFTALSDIEVIKHDVEGAFYHNMNMLEDSGRSLAEVAT 218
Qy 395 YLSEFLFGTEVVRRAQANALATPDHAGILKSLDXKYDYLPS-----LDRVVLAM 449
Db 219 TRPEFMFGDV-----VVAVNP-----DPRYDLGLGNVILPIANLILPIV 259
Qy 450 LD-----IVRRIRPADKLPIDFKVIDKKFKGPTKKKADVPFDSKVPYSDK 496
Db 260 GDEHADPELGTGVYKITPAHDPNDPLVGORHNLPGVNNMNDGTMNELAFESGH---DR 316
Qy 497 FHAMLSKDKKEFAKAIKKDPAVELSKSVIAAARA-----IOAD 535
Db 317 FEAR-----KAVVAKLEEFIGALVYIEKRVHSGHSERTGVVPEPLSTQWFKMDQLAKN 371
Qy 536 AMANVAIEKGRKRLFFALREMYPERALPS-----DANFTMRMSYGSIKGEPDGS 586
Db 372 AIANDTEDEK-----VEFYPRFNDTFLQWENHVDVISHQLMWGH-----QIP 416
Qy 587 AMYNYHTTGKGVLEKODEPKSDEFVAQENILD 617
Db 417 AMYN--ADGEMVYGEAPEGSGWQDEVDLD 445

RESULT 14

DNA helicase (EC 3.6.1.-) MTH1770 [validated] - Methanobacterium thermoautotrophicum (str
N:Alternans: adenosinetriphosphatase (EC 3.6.1.3)
C:Species: Methanobacterium thermoautotrophicum
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 26-May-2000
C:Accession: D69103
R:Smith, D.R.; Doucette-Stamm, L.A.; Delonghery, C.; Lee, H.; Dubois, J.; Aldredge, T.;
Qiu, D.; Spadafora, R.; Vicalaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jivanli, N.
Kl, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7153, 1997
A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funct
A:Reference number: A69000; MUID:98037514
A:Accession: D69103
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-666 <MTH>
A:Cross-references: GB:AE000932; GB:AE000666; NID:92622894; PIDN:AA8636.1; PID:9262290
A:Experimental source: strain Delta H
C:Genetics:
A:Gene: MTH1770
C:Complex: dodecamer; double hexamer [validated, MUID:20144074]
C:Function: <HEL>
A:Description: EC 3.6.1.-; DNA helicase; ATP-dependent DNA unwinding activity [validated
C:Function: <ATP>
A:Description: EC 3.6.1.3 [validated, MUID:20144074]; adenosinetriphosphatase, DNA-deper
C:Function: <DNA>
A:Description: ATP-independent single-stranded DNA-binding; ATP-dependent double-strande
C:Superfamily: Archaeoglobus replication licensing factor MCM-type homolog; MCM homolog
C:Keywords: cell cycle control; DNA replication initiation; hydrolase
F:98-570/Domain: MCM homolog <MCM>

Query Match 3.2%; Score 120; DB 1; Length 666;
Best Local Similarity 19.3%; Pred. No. 6.6;
Matches 116; Conservative 94; Mismatches 232; Indels 158; Gaps 28;

Qy 119 LPIPLSVKYLK-----YKVTDKVSGQL-----KGIDEMERLKAQEV 159
Db 96 IPIRLRSKIGKGFVAVDGIYKRTDEIRPIYKAVFECRCMRHNAVOSTMITE-PSL 154
Qy 160 COELA-----KKNADENQLOCIIVPF--YSNNEFLIYDYFKDVRYVFAFPSSVG 208
Db 155 CSECGRSRFLQDSEFLDTQTLKLEPLENLSGGEOPROQTIVVLELDLVTLTPGDIY 214
Qy 209 KFGGDT-----NNMMWPRHT-----GDFSVFRYACADNRPAYSKDNKPY-RPV 252
Db 215 RVTGLKRVYRDERTKRFKNFTIGNTFLEQEFBELQISEDEEKIKELAGDPNIYKII 274

Qy 253 YFAVSMOGYKADYAMTIGEPGSTDRYLTSWGEDRIENENNRIEYGIKQIMWEAM 312
Db 275 RSTAPSHIGYREVEALALQLFGGKEL-----DDKTRLNGDIHILIV-DPGIGSKM 328
Qy 313 SADOATRIKYSKYAQSNYK-----NSIGMRGLARLDVYGRKRAERAPADIRKNGK 368
Db 329 -----LKVYSKLAPRGITYSKGTSQVGLTAAVROE-----FGGSLG- 369
Qy 369 SAYVGD-----VLSLEKAYKEGAKANREMYLSETFEGTEVVRFAQANALATNPDAHA 424
Db 370 ALVLDKKNVYVDELDRKREDRSHIE-----ALEQGITSTAKA 409
Qy 425 GILKSLD-----DKYKDYLSLDRKVLPA-----MLDIVRRIRPADK 461
Db 410 GIMATLNSRCGVLAANPKFGFRPSYKSIABQID--LPSTILSRFLDIFVEEDKPDPEK 466
Qy 462 LPDLFKNYIDKKFGDPTKKYADFPDKSVVYRSKFAHMLKSMOKERPAKAIENDPAVEL 521
Db 467 DRELARHIL-KTHKED--HMFELDPELL--RKYIAYAR---KVRPVLIDEMAKOV 514
Qy 522 SKSVIAAARAIQAD---AMANVAIEKGRKRLFFAGLEMYPGRALPSDANFTMRMSYGS 577
Db 515 LEDTVYSMRASADSDSVPTTAQLEALVRLSEASAKIKIKHEVNEADAKAKIKLSQAC 574
Qy 578 IK--GYEPQDGMVNYHTTGKGVLEK---QDPKS--DEFAVOENILDLEFRKNYGRVAEN 630
Db 575 LKQGYDPE-----TGKIDIDKVEGRTPKPSBROKFRLLLEIKYEEDDYGGRAPTN 625

RESULT 15

translation elongation factor EF-G VC0361 [Imported] - Vibrio cholerae (strain N16961
C82332
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: C82332
R:Heidelber, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gattm, M.L.; Dodson, R.
charidson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers
1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833
A:Accession: C82332
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-698 <HEI>
A:Cross-references: GB:AE004124; GB:AE003852; NID:96547770; PIDN:AAF93534.1; GSPDB:GN
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC0361
A:Map position: 1
C:Superfamily: translation elongation factor G; translation elongation factor Tu homo
F:11-145/Domain: translation elongation factor Tu homolog <ETU>

Query Match 3.2%; Score 118; DB 2; Length 698;
Best Local Similarity 18.4%; Pred. No. 9.5;
Matches 120; Conservative 83; Mismatches 202; Indels 246; Gaps 26;

Qy 8 ILLGALLL-GASGVAKADKGMW-----LNLNELQENLDRRRELGTPLDLSL 54
Db 105 VLDAVAVVFCGTSGEVPEPSETVMQADKYGVPRMVFVKMRACADFLRYYG----- 156
Qy 55 YSPDKPSIANAVI-----FGGCGTGITV-----SDQGLIFT----- 86
Db 157 -QIKHRLGANPVIQOLNIGAESEEGVYIDLKMKALNNNADQGSFYEEIPADMLELA 215
Qy 87 ---NHHCYGAIGOSYVDHYLDGFGS-----RTMGELPPIGLSVYLYKRIY 133
Db 216 QEMRNHILVEAAAESEEMERYLELLEDGELSEYKIQALRRTINNETIVLAACGSAFKNGV 275
Qy 134 K-VTDKY-----EGOLKGITDEMERLKAQAEVCOELAKKENADENQLOCIPEPFSN 183
Db 276 QAVLDAVLEFLPSPIDVAIKGIDD-----RENSEVERHADN-----EPF--S 316

```

QY 184 NEYFLIVYDFKDVBMVFAFPSSVKGFGDPTDMMMPRHRTGDSVFVRVYAGADNRPPEYS 243
Db 317 SLAFKTIATDF-----VGSUFTIRVYSGVYNSGDVY 348
QY 244 KDNKPYKPYEAAYSOGYKAD-----DYAMTIG----- 272
Db 349 NSYKOKKEREGRIVQMHANKRDEIKERAGDIAAIGLKDVTTGDTLCPDNHVVILERME 408
QY 273 FPGSTRYLTSWGVEDRIENENNRIEVRGIRKGIKWEKMSA--DOATRIYASKYAS-- 329
Db 409 EP-----EPVIOIAVEPRSKADOEKMGIALGKLAEDPSFRVETDAETGOTLI 456
QY 330 -----ANYMKNSIGMNRGLARLDVIGRK--RAEERAFADWIRKNGKSAVYG--- 373
Db 457 SCMGELHLDIYDRMKRREGVDCNVGKPOVAYARETIRGKSEVGKFFVQSGRGQYGHVW 516
QY 374 -----DVLSSLEKAYKEGAKANREMTYLSSETLFGGT 404
Db 517 LKIEPAEPGQGFVFDVADIAGVLPKEFINPVAKIEEQMNNNGVLAGYPVLDVAKATLFDGS 576
QY 405 -EYVREAOQFANALATNPDAHAGILKSL-----DDKYKDYLPsIDRK--VL 446
Db 577 FHDVDSSEMAFKTAGSMAFRKGALEAPVLLPELPMKVEITTPEDMMGDVVGDLNRRGII 636
QY 447 -----PAMLDIVRRRIPADKLDPDIFKNVIDRK-----FKGDTKKKYAD 483
Db 637 EGMDEGPAGIKITHAKVP---LSEMFGYATDLRSATQGRASYSMEFAEYAD 684

```

Search completed: October 17, 2002, 22:40:19
 Job time : 49 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 17, 2002, 21:37:45 ; Search time 25 Seconds

(without alignments)
1102.733 Million cell updates/sec

Title: US-10-008-355-2

Perfect score: 3719

Sequence: 1 MQMKLKSILGALLGASG.....LFMDKMGCCPRILQELKLI 712

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	123.5	3.3	776	YL44_CAEEL	Q05036 caenorhabd
2	122	3.3	1104	COLA_CLOPE	P43153 clostridium
3	121	3.3	1628	NAGH_CLOPE	P26831 clostridium
4	116	3.1	952	TOP1_SRRCO	Q9X909 streptomyces
5	114	3.1	620	DNK_PORPU	P30723 porphyra pu
6	114	3.1	734	GLGB_AGR7	P52579 agrobacteri
7	114	3.1	747	VIVC_BPT7	P03725 bacterioph
8	113	3.0	1047	RIRI_CHIMU	Q9P193 chlamydia m
9	112.5	3.0	657	CN16_HAEIN	P44764 haemophilus
10	112	3.0	556	EST2_CAEEL	Q07085 caenorhabd
11	112	3.0	1504	DPO2_YEAST	P14284 saccharomyc
12	111.5	3.0	878	ACOM_RICPR	Q9ZC14 rickettsia
13	111.5	3.0	922	YK6_CAEEL	P42173 caenorhabd
14	111.5	3.0	2410	MOKI_SCHPO	Q9U8K8 schizosacch
15	110.5	3.0	1039	SYI_MERYA	Q58357 methanococ
16	109.5	2.9	768	PARC_NEIGO	P48374 neisseria g
17	109	2.9	681	MP10_HUMAN	Q00566 homo sapien
18	109	2.9	3672	LMU2_CAEEL	Q21313 caenorhabd
19	108.5	2.9	507	YV42_CAEEL	Q18416 caenorhabd
20	108	2.9	638	NEC2_PIG	Q03333 sus scrofa
21	108	2.9	1341	RPAL_METJA	Q08445 methanococ
22	106.5	2.9	747	Y030_UREPA	Q9P4B5 ureaplasma
23	106	2.9	1592	GTF2_SRRDO	P27470 streptococ
24	105	2.8	1023	HLV1_ECOLI	P09983 escherichia
25	105	2.8	1391	RPOB_MTCPN	P78013 mycoplasma
26	104	2.8	616	ACOC_SOLTU	Q04916 solanum tub
27	104	2.8	908	H104_YEAST	P31539 saccharomyc
28	104	2.8	1060	DP3A_LACLA	Q9C170 lactococcus
29	104	2.8	4563	APA_HUMAN	P44114 homo sapien
30	103	2.8	320	PTA_MYGE	P47541 mycoplasma
31	103	2.8	843	PUL4_THEMA	Q33840 thermotoga
32	103	2.8	1131	APCE_ANASP	P80559 anaerobac
33	102.5	2.8	396	GTS1_YEAST	P40956 saccharomyc

34	102.5	2.8	718	1	FLGE_HELPY	Q9ZKY0 helicobacte
35	102.5	2.8	798	1	UNR_RAT	P18295 rattus norv
36	102	2.7	555	1	MCRA_METPE	P12971 methanother
37	102	2.7	560	1	YDEN_ECOLI	P77318 escherichia
38	102	2.7	626	1	HTPG_BACSU	P46208 bacillus su
39	102	2.7	656	1	HTPG_MYCLE	Q03612 mycobacteri
40	102	2.7	742	1	KM65_YEAST	Q03566 saccharomyc
41	101.5	2.7	627	1	CACP_CANTR	Q00614 candida tro
42	101.5	2.7	629	1	KSK_MOUSE	P48025 mus musculu
43	101.5	2.7	675	1	RG59_MOUSE	Q54828 mus musculu
44	101.5	2.7	770	1	DBP4_YEAST	P20448 saccharomyc
45	101	2.7	500	1	Y039_BORBU	O51068 borrelia bu

ALIGNMENTS

RESULT 1	ID	YL44_CAEEL	STANDARD:	PRT:	776 AA.
AC	005036;				
DT	01-FEB-1994 (Rel. 28, Created)				
DT	01-FEB-1994 (Rel. 28, Last sequence update)				
DT	01-MAR-2002 (Rel. 41, Last annotation update)				
DE	Hypothetical 86.9 kDa protein C30C11.4 in chromosome III.				
GN	C30C11.4.				
OS	Caenorhabditis elegans.				
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;				
OC	Rhabditidae; Peloderinae; Caenorhabditis.				
OX	NCBI_TaxID=6239;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=BRISTOL N2.				
RX	MEDLINE=94150718; PubMed=7906398;				
RA	Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,				
RA	Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,				
RA	Craxton M., Dear S., Du Z., Durbin R., Favell A., Fraser A.,				
RA	Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,				
RA	Johnston L., Jones M., Kersey J., Kristen J., Laister N.,				
RA	Latille P., Lightning J., Lloyd C., Mottmore B., O'Callaghan M.,				
RA	Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Shownkeen R.,				
RA	Sims M., Smalton N., Smith A., Smith M., Sonhammer E., Staden R.,				
RA	Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,				
RA	Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,				
RA	Wohlman P.,				
RT	"2.2 Mb of contiguous nucleotide sequence from chromosome III of C.				
RL	"elegans.";				
RL	Nature 368:32-38(1994).				
CC	-1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY. STRONG,				
CC	TO YEAST MS13.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement. (See http://www.isb-sdb.ch/announce/				
CC	or send an email to license@sdb-sdb.ch).				
CC	-----				
DR	EMBL, L09634; AAA27967.1; -				
DR	PIR, S44784; S44784.				
DR	HSSP, P19120; INGC.				
DR	WormBep, C30C11.4; CRE00103.				
DR	InterPro, IPR001023; HSP70.				
DR	PIfam, PF00012; HSP70.1.				
DR	PROSITE, PS00297; HSP70.1; FALSE_NEG.				
DR	PROSITE, PS00329; HSP70.2; 1.				
DR	PROSITE, PS01036; HSP70.3; 1.				
KW	Hypothetical protein: ATP-binding; Multigene family.				
SQ	SEQUENCE 776 AA; 86896 MW; 20FA975FB649FA9C CR664;				
Query Match	3.3%;	Score 123.5;	DB 1;	Length 776;	
Best Local Similarity	21.1%;	Pred. No. 2.2;			


```
DB 713 KDMNSK---LNDILKELSKSNNGYKTVATAFVNHKVDNGNVYDVEFGHMT----- 763
QY 413 ANALANPNPAHA-----GLKSLDCKYKDYLSLDRKYVPAALDIYRRKIPADKLPDIRK 467
DB 764 ---DNTNTVHVHVKRPAVKS-----DSSV-----IYEEBIN-----PD 794
QY 468 NVIDRFKFGDTKKY-----ADFEVDEKSVYPSDKFHAMLSMD-----K 506
DB 795 GTSKSEDEEIKAYEMWDFGDEKSNKAKTHTKYNK-----TGEYEKLYVTDNNNGSINTGE 849
QY 507 EKFPAKAEKDP-----AVELSKSVIAAARAIGADAMANAAYAEKGR-- 548
DB 850 SKKIKVYEDKPEVINESEPNNDFEKANQIAKSNMLVKGTLSBEDYSDKYFDVAKGV 909
QY 549 -----LFFGLREMY-----PGRALPDAFTMASTGSIKGYEPDQG 586
DB 910 KITLNLNLSVGIWTWLYKKGEDLNINYLATGNDGIVLKKEKTELPGRYLYSVTYDNOQG 969
QY 587 AMYNYHTTGKGVLEKODPKSDPEAVQENILDFRTKNYGRVAENQGLHIAFLSNNDITGG 646
DB 970 ---TIVVNVKGNLKNVEKETAADAKE-----VEN-----NND----- 999
QY 647 NGSGPYEDK-----NGRLIG-LAPDGNWEAMSGDIEFEPDQRTI---SYDIRVYLEM 695
DB 1000 -----FDRAMKVDNSKIVGTLSNDLDKDIYSIDIQNSDLNIVENDNIKMMLLVS 1053
QY 696 ID 697
DB 1054 AD 1055

RESULT 3
MAGH_CLOPE
ID MAGH_CLOPE STANDARD: PRT: 1628 AA.
AC P26831:
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Hyaluronoglucosaminidase precursor (EC 3.2.1.35) (Hyaluronidase)
DE (Mu toxin).
GN MAGH OR CPE0191.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1502;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=13 / Type A;
RX PubMed-11792842;
RA Shimizu T., Ohnishi K., Hirakawa H., Ohshima K., Yamashita A.,
RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT "Complete genome sequence of Clostridium perfringens, an anaerobic
RT flesh-eater."
RL Mol. Gen. Genet. 243:215-224(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=13 / Type A;
RX PubMed-11792842;
RA Shimizu T., Ohnishi K., Hirakawa H., Ohshima K., Yamashita A.,
RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT "Complete genome sequence of Clostridium perfringens, an anaerobic
RT flesh-eater."
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
CC -1- FUNCTION: Putative virulence factor which is likely to act on
CC connective tissue during gas gangrene.
CC -1- CATALYTIC ACTIVITY: Random hydrolysis of 1,4-linkages between N-
CC acetyl-beta-D-glucosamine and D-glucuronate residues in
CC hyaluronate.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- CAUTION: The partially purified protein from strain CPN50 is
CC approximately 70 kDa smaller than the sequence indicated here.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

```
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb.ch).
CC -----
DR EMBL: M81878; AAA23259.1; -.
DR EMBL: AP003185; BAB79897.1; -.
DR PIR: S27540; S27540.
DR InterPro: IPR00421; FA58_C.
DR Pfam: PF00754; P5_P8_type_C; 1.
DR SMART: SM00231; FA58C; 1.
KW Hydrolase; Glycosidase; Toxin; Repeat; Signal; Complete proteome.
FT SIGNAL 1
FT CHAIN 31 1628
FT DOMAIN 703 1032
FT REPEAT 703 748
FT REPEAT 848 894
FT REPEAT 987 1032
FT REPEAT 147 147
FT VARIANT 172 175 G->A (IN STRAIN CPN50).
FT VARIANT 250 250 V->M (IN STRAIN CPN50).
FT VARIANT 548 548 A->E (IN STRAIN CPN50).
FT VARIANT 558 558 D->E (IN STRAIN CPN50).
FT VARIANT 614 614 G->S (IN STRAIN CPN50).
FT VARIANT 944 944 I->V (IN STRAIN CPN50).
FT VARIANT 950 950 N->S (IN STRAIN CPN50).
FT VARIANT 979 979 T->I (IN STRAIN CPN50).
FT VARIANT 982 982 I->L (IN STRAIN CPN50).
FT VARIANT 1042 1042 I->F (IN STRAIN CPN50).
FT VARIANT 1043 1628 MISSING (IN STRAIN CPN50).
SQ SEQUENCE 1628 AA; 182474 MW; D4252A2512BEB69 CRC64;

Query Match 3.3%; Score 121; DB 1; Length 1628;
Best Local Similarity 20.1%; Pred. No. 8.6;
Matches 129; Conservative 78; Mismatches 209; Indels 226; Gaps 36;

QY 176 IYEPFYSNN-----EFLIYDVFKDYRMVFAPSSGKRGCDTMMMPRHGDSVFPV 231
DB 111 VVDNFFNKKIPHDESF-----FDEKMDANIVSKDGYVIGEDTDAFY---GVTLKHV 163
QY 232 YGADNRPAEYSGKDNKPYKPYFAAVSMQGYKADYAMTIGPGSDYRL-TSMGVEDRI 290
DB 164 F---NQLEEGNK-----IOSFRADYA-EVAIRGFIEGYGPNWSMEDA 204
QY 291 E-----NENNPRIEVRIGIKGIMKEAMSADATRIKYASKYQSANWK 334
DB 205 ELMKFGDYKLNQYVAPKDDP-----YHNSKWRDLYPEEKLSKIK---KLAQVGNETK 255
QY 335 NSI-----GNRGLA-----RLDYIGRK---RAEERARA----- 360
DB 256 NRYVALHPFMNPNVFEDEENYQNDLVGIKAKFTQLENDVROFALILADDAAPAQAGS 315
QY 361 -----DWIRKNGSAVYGDVLSL-----EKAYKEGAKA-----NREMYLSETLF 401
DB 316 MYVKLLTDLTRLWE--QOSTYPLDKTDLMPCRPSDYTGSGSSNQLKELKADNVSIVVT 373
QY 402 GGT---EYVRFQAFANALATN--PDHAG-----ILKSLDCKYKDYL-----PS 440
DB 374 GGRIMEYDE--NFANFMNNISTEGHPGRAPPFWMNCPNSKQHLIMGNDYFLHPG 431
QY 441 LDRKVLPAALDIYRRIRPADK-----LPDIFKNVIDKKRKGDTK-----KYADF---VF 486
DB 432 VPSKIDGIVLPMDDAEKNSALFAIDYANVINWKEADENMDSFKYDHGTAETT 491
QY 487 DKSVPYPSDKFAMLSKMD--KEKFAKIEKDPAYELSVIAAARAIGADAM---ANAY 541
DB 492 NSIALRELSKIMINOMDGRVPLQESVELAPKLEAFQOXDGSASIEDALDELAEFT 551
QY 542 ALEKGRLEFAGIREMYPGRALPSDA-----NFTMMSYGSIKG-----YEPDQGA 588
DB 552 NIQKA-----ADYKKNNPNERTDQIIVLNCWEDTMDAIGYIKSAIAIEEGDEAAM 606
```

```
OY 589 YNYHTGKGVLEKODPKSDEFAVQENIIDLFRTKNYG-----RYAENGQLHT----- 635
DB 607 ANY-SEAGAFERK-----SKTYGFHYVDHTETAEVGVQHIYPIKLS 646
OY 636 -----AFLSN-NDITGNSGSPYFDKN 636
DB 647 MGNLTSVIGSTIVDPNRILATYISNRDAPGPNPDN-IFDNN 687

RESULT 4
TOP1_STRCO
ID TOP1_STRCO STANDARD: PRT: 952 AA.
AC 09X909;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE DNA topoisomerase I (EC 5.99.1.2) (Omega-protein) (Relaxing enzyme)
DE (unwisting enzyme) (Swivelase).
GN TOPA OR SCH5.06C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycetales; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=A3(2);
RA Seeger K.J., Harris D., Bentley S.D., Parkhill J., Barrett B.G.,
RA Rajadream M.A.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: THE REACTION CATALYZED BY TOPOISOMERASES LEADS TO THE
CC CONVERSION OF ONE TOPOLOGICAL ISOMER OF DNA TO ANOTHER.
CC -1- CATALYTIC ACTIVITY: ATP-independent breakage of single-stranded
CC DNA, followed by passage and rejoining.
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- MISCELLANEOUS: WHEN A TOPOISOMERASE TRANSIENTLY BREAKS A DNA
CC BACKBONE BOND, IT SIMULTANEOUSLY FORMS A PROTEIN-DNA LINK, IN
CC WHICH A TYROSYL OXYGEN IN THE ENZYME IS JOINED TO A DNA PHOSPHORUS
CC AT ONE END OF THE ENZYME-SEVERED DNA STRAND.
CC -1- SIMILARITY: BELONGS TO PROKARYOTIC TYPE I/III TOPOISOMERASE
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL035636; CAB38480.1; -.
DR HSSP; P06612; 1ECL.
DR InterPro; IPR003601; DNATOP1_ATP_bind.
DR InterPro; IPR003602; DNATOP1_DNA_bind.
DR InterPro; IPR000380; Pro_topoisomerase.
DR InterPro; IPR002936; Toprim.
DR Pfam; PF01131; Topoisom_bac; 1.
DR Pfam; PF01751; Toprim; 1.
DR PRINTS; PR00417; PRTPISMRASEI.
DR SMART; SM00437; TOP1AC; 1.
DR SMART; SM00436; TOP1BC; 1.
DR SMART; SM00493; TOP1IM; 1.
DR PROSITE; PS00396; TOPOISOMERASE_I_PROK; 1.
KW Isomerase; Topoisomerase; DNA-binding.
FT ACI_SITE 334 334 DNA CLEAVAGE (BY SIMILARITY).
FT SEQUENCE 952 AA; 103584 MW; 1C98DABD7A629295 CRC64;

Query Match 3.1%, Score 116, DB 1, Length 952;
Best Local Similarity 18.8%, Pred. No. 8.6;
Matches 124; Conservative 96; Mismatches 279; Indels 162; Gaps 29;
```

```
OY 156 AOEVCQE-----LAKKENADENOLC-----IVEPYSNNEYFLIYVDKX-VRKVEAPP 204
DB 89 LKDLKESDELFLATVDEDEREGBALAMHQLQEVLRKPIPKR--WVFHEITKQAIKRAAVANP 146
OY 205 SSVCK-----FGGDTDNMMPRHHGDSVFVYVYAGADNRPAPEYKDKPKY 249
DB 147 RELNOKLVDQAQETRRILDRLYGVEVSPVLMKVRPLRSAGVQSVATYTLVVERERETAF 206
OY 250 KPYVEFAAVSMQGYKADYVAMTIGPPGSDRYLTSGVGEDRIENENNPRIEVRGIKOGIMK 309
DB 207 KSAET--WDLTGTPATGRAGASDPSSLVRLQT-----VDRRAVQG--- 247
OY 310 EAMSAQATRIKIVASK-YAOSANYWKNISGNKGLARLDVIGKRAEERAPADW----- 362
DB 248 --RDDDSLGQLKSNANTLHLDENANRALAALLENTRFANRVSQSPYRRSPAPARTTLQ 305
OY 363 ---IRKNGKSAVYGDVUSLSEKATKEG--AKANREMYLSETLFGTEV-VRFQFANAL 416
DB 306 QEASRKLGFGA--KSTMQVQKLYENGITYTMRDSTLSDTAVSAAARQVTOLYGADYL 363
OY 417 ATNPDAHGLIKSLDDKXKDYLDPLSDRKVLPA-----MLDIYRRRIAPDKLPDI 466
DB 364 PPQPTTYAGKVKNAQEAHEAIRPSGDRRTPAETGTLGDDFKLYELMKRTVASOMKDAT 423
OY 467 KNVIDKRFKGGDTKKYADVFEDKSVVPYSDKFFHMLKSM-----DKERFAKAI- 513
DB 424 GNSVYTKIGGAASGRDVEFASCKTIT--FHGFLKAAVEGADDPNMLDREHRLQVYA 481
OY 514 -----EKDPAVELSKSYI-----AAARAIQADAMANAVALKEG 546
DB 482 EGDALTAETIVDGHATPARYTEASLYKELEREIGRPSTVSIIGTLIDRGVFKG 541
OY 547 KRLF-----FA--GLREYPGRALPSDANFTMRM--SYGSIKGYEPODGANYNH----- 592
DB 542 TALVPSTLSFAVNVNLEHNFRLV--DVFETARMEDLDLRARGAQSVPPLRRPYGEG 599
OY 593 --TGKGVLEKODPKSDEF-AVQENIIDL-----FRTANGRYANG 631
DB 600 DGTGGCGAADNGNDGHLGLGKELVTDLGAIDAREVSSFPVGNIDKIRVGRYPYVERG 659

OY 632 Q 632
DB 660 E 660

RESULT 5
DNAK_PORPU
ID DNAK_PORPU STANDARD: PRT: 620 AA.
AC P30723;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Chlorophyll a protein (HSP70).
DE Protein.
DE Porphyra purpurea, and
DE Porphyra umbilicalis (Laver).
OS Chloroplast
OC Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae; Porphyra.
OX NCBI_TaxID=2787, 2786;
RN [1]
RP SEQUENCE FROM N.A.
RP SPECIES=P.purpurea; STRAIN=AVONPORT;
RA Reith M.E., Munholland J.;
RT "Complete nucleotide sequence of the Porphyra purpurea chloroplast
RT genome.";
RL Plant Mol. Biol. Rep. 13:333-335(1995).
RN [2]
RP SEQUENCE FROM N.A.
RP SPECIES=P.umbilicalis; STRAIN=AVONPORT;
RA MEDLINE=92077124; PubMed=1720741;
RA Reith M., Munholland J.;
```

RT "An hsp70 homolog is encoded on the plastid genome of the red alga,
 RT Porphyra umbilicalis."
 RL FEBS Lett. 294:116-120(1991).
 CC -1- FUNCTION: ACTS AS A CHAPERONE (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: U38804; AAC08201.1; -
 DR EMBL: X62240; CA44160.1; -
 DR PIR: S19660; S19660.
 DR HSSP: P04475; IDG4.
 DR InterPro: IPR001023; HSP70.
 DR Pfam: PF00012; HSP70; 1.
 DR PROSITE: PS00297; HSP70_1; 1.
 DR PROSITE: PS00329; HSP70_2; 1.
 DR PROSITE: PS01036; HSP70_3; 1.
 DR Chapterone; ATP-binding; Heat shock; Chloroplast.
 KW SEQUENCE 620 AA; 67619 MW; C9FB4713C142FECF CRC64;
 SQ
 Query Match 3.1%; Score 114; DB 1; Length 620;
 Best Local Similarity 19.8%; Pred. No. 6.4; 240; Indels 198; Gaps 32;
 Matches 18; Conservative 82; Mismatches 198; Gaps 32;
 QY 51 LDSLVSFDPKPSINAVVIFGGCTGTVSPDGLIFTNHHCY-----GAIQOSQTVYD 102
 DB 1 MGRVVGIDIGTNSVIAVMGGRPTVIPNAEGPRTASVAAYIKSGDKLVGLQARQAVIN 60
 QY 103 HD---YLRDGFVRFMGE--ELPIRGLSVYKRYKRYKTDKYGOLKGTIDMERIRK- 155
 DB 61 PENTYFVYKFFIKRQNEISQEIROTSYNVTGSSSIKI-----BCPALMKD 107
 QY 156 --AOEVOCELAKK--ENADE-----NOLCIPEFYSNNRYFLIYVDVK---DY-RMV 200
 DB 108 FAEELISAQVLRKLVEDASTYLGELYTQAVITPAIFNDORATDACKIAGLIDVLRIT 167
 QY 201 FAPPSVGRKGGDTN-----MMMPRHGTDFSVFRVYAGADNRPAEYSKDNKPYKVPYF 254
 DB 168 NEPTAASLSTYGLDKONNETILVFDLGGGTFDVSILEYGDV-----F 209
 QY 255 AAVSMOG--YKADYVAMTIGFPGSTDRYLTSGVEDRIENENNRPIEVAGTIGWKEA 311
 DB 210 EVLTSGDTHLGGDFF-----DOQIVEMLIKDFKQSE-----GIDLGXKDRQA 251
 QY 312 MSADQATRIYASKYA--OSANTYKNSICGNRGLARLDVIARRAERAFADMTIRKNGKS 369
 DB 252 LQ-----RLTEASEKAKIELSNLTQTEINLPFTATOD--GPRHLE-----KTVTR 295
 QY 370 AAVGDVLSL-----EKAYKEGAKANREMYLSETLP--GGEVYVFAQAPANAL--- 416
 DB 296 AKREELCSRLIDKCSIPVNNALMD--AKLEASSIDEVVLVGSGSTRIPALQGWKRLIGK 352
 QY 417 ----ATNPD-----AHAGILKSLDDKYYKDYLPISLDRKVLPAMLDIYRRRIIPADKLPD 464
 DB 353 DPQNSVDPDEVAIGAAGVGLAG--EVKDLI-----LIDVTPLSLGVETLGG 399
 QY 465 IFKNVNDIKRKGGDKTKKADVPFDKSVVPYSDKTHAMLSMDKEKFAAIKDAVELESKS 524
 DB 400 VMKTIIPRNTTIPKK-----SEVFSTAVNQPWEL--Q 432
 QY 525 VIAAARAIQADAMANAIAIEKGRLEFAGLRMPG--RALPS-----DANFTMRSYG 576
 DB 433 VLQGERELTMD-----NKSLSGTFRLDGIIMPARAGVQJLEVTVIDIDANGL----- 477
 QY 577 SIKGYEPDQAMTYNTHTGKGVLEKODPKSDEFAVOENT-LDLFRTKN 623
 DB 478 SYRAKEKATGEOSTITSGASTLPKDDVERHVMKAEENFNDVQDKRRKN 525

RESULT 6
 GLGB_AGRU
 ID GLGB_AGRU STANDARD; PRT; 734 AA.
 AC P52979;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE 1,4-alpha-glucan branching enzyme (EC 2.4.1.18) (glycogen branching
 DE enzyme).
 GN GLGB.
 OS Agrobacterium tumefaciens.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Rhizobium.
 OX NCBI_TaxID=358;
 RX MEDLINE=99069330; PubMed=9851999;
 RP STRAIN=A348;
 RC MEDLINE=99069330; PubMed=9851999;
 RA Ugade J.E., Lepek V., Uttaro A.D., Estrella J., Iglesias A.,
 RA Ugade R.A.;
 RT "Gene organization and transcription analysis of the Agrobacterium
 RT tumefaciens glycogen (glg) operon: two transcripts for the single
 RT phosphoglucomutase gene."
 RL J. Bacteriol. 180:6557-6564(1998).
 CC -1- CATALYTIC ACTIVITY: Formation of 1,6-glucosidic linkages of
 CC glycogen.
 CC -1- PATHWAY: THIRD STEP IN GLYCOGEN BIOSYNTHESIS.
 CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO
 CC KNOWN AS THE ALPHA-AMYLASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: AF033856; AAD03472.1; -
 DR InterPro: IPR000461; Alpha_amylase.
 DR InterPro: IPR004193; Isoamylase_N.
 DR Pfam: PF00128; alpha-amylase_1.
 DR Pfam: PF02922; isoamylase_N_1.
 KW Glycogen biosynthesis; transferase; Glycosyltransferase.
 FT ACT_SITE 417 417
 FT ACT_SITE 470 470
 FT ACT_SITE 538 538
 FT BY SIMILARITY.
 SQ SEQUENCE 734 AA; 83623 MW; 70A3CD5A77F31E6 CRC64;
 Query Match 3.1%; Score 114; DB 1; Length 734;
 Best Local Similarity 20.1%; Pred. No. 8.1;
 Matches 154; Conservative 84; Mismatches 261; Indels 266; Gaps 43;
 QY 108 DGEVSRFM---GHELRPIGLSVYKRYKRYKTVKVGOLKGTIDMERIRKAEVQDELA 164
 DB 43 EGSARCFIPGAEEVSVLTLDGNFVGLKQIDP--DGFEGRIDLSRKQRYRRYK- 97
 QY 165 KRENADENOLCIYEPFYSN-----NEYFLIYDV-----FKDVR---MVF 201
 DB 98 ----DDAEVAIVDPYISFGVLPMDYFVBESISGYSGMARIPLKLEGGVGFPAW 152
 QY 202 APPSSVGRKGGDTNMMMPRH-----TGDFSVFR--VYAGADNRPAEYSKDNKPYKPV 252
 DB 153 APNGRRVSVGDDNNMDGRHVRFRKDTGIMELFAPADVACA-----YK-- 197
 QY 253 YFAAVSMOG--YKADYA-----WTIGFPGSTDRYLTSGVEDRIENENNRPIE 299
 DB 198 -FEILGANGELPLKADPYARRKELPRKNASVTAPELTQW-----EDQAHREHMAQVD 250
 QY 300 VR-----GIKGIWKEA-----MSAD--QATRIYASK----- 325

```

Db 251 QRRPISIEYVHAGSMOSESDGTFLSMDELAEQLIPYCGDMGFTHEIFLPTHEHPDPSW 310
Oy 326 -----YAOSANYWKNKNSGNRGLARDLYGKRAERAPADWL----- 363
Db 311 GYOTTGLAAPTARF-----GDEGEFARF-VNCAHKVYGIGVLLDWPAPHPEDHGLRMFD 364
Oy 364 -----RKNGSAVYG-----DVLSSL-----EKAYKSGAK--ANREMTY 395
Db 365 GTALTEHADPRGCFHPDNTATITNGRIEVMKSYLLNNALYMAEKHLDGLRYDAVASMLY 424
Oy 366 LSETL-----FGC-----TEVVFAPAFANALATNPDAHAGILKSLDKKYDLPISDR 443
Db 425 LDYSRKEGEMIPNEYGRENLESVRFLOKMSLTVYG--THPGVMTIAESTS--WPKYSQ 480
Oy 444 KVLPLMLDIYRRIRPADLPLDIFKNVIDKKFGDTKKADPEYEDKSVYPSKFTAM--- 500
Db 481 PVHEGGLG-----FGFKMNMGFMDLSTY---FSRE--PVHKKFHHOELT 520
Oy 501 --LKSMDKEKFAKIEKDPAVELSKSVIAAARAIOADMANAYAT-----EKGKRLFFA 552
Db 521 FCLLAFTENFVLPISHDEVHNGKSLIAKMSGDMOKFANLRSTYGFPMGYPGKLLFM 580
Oy 553 G-----LRE--MYPG-RALPSDANFTMRMSYG-SIKGYEPDQGAN-- 588
Db 581 GQEFQAWSEWSEKSGSLDWNLRQYPMHEGMRRLVRLNLTYSKALHARDEPQDFRMLY 640
Oy 589 YVYHTTGKGVLEKOPKSDSEFAVOENIDLFTKRY-----GRAENQGLHIALNSND 642
Db 641 VDDHNSVFAWLRTAPGEKPAVANICNLTPVRYENYVPLGVAGNRE-----ILMTD 692
Oy 643 --ITGNSGSPVFDKNGRLIGLAFDGNMEAM-----SGDIEEPD 680
Db 693 AEITGG-SGK---GNGGRQAQVADAGEIGAMLVLPPLATIMEPE 733

```

```

RESULT 7
VIVC_BP77 STANDARD: PRT: 747 AA.
AC P03725:
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-MAR-1989 (Rel. 10, Last annotation update)
DE Internal virion protein C.
GN 15.
OS Bacteriophage T7.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae;
OC T7-like phages.
OX NCBI_TaxID=10760;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=83241725; PubMed=6864790;
RA Dunn J.J., Studier F.W.;
RT "Complete nucleotide sequence of bacteriophage T7 DNA and the
RT locations of T7 genetic elements."
RL J. Mol. Biol. 166:477-535(1983).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; V01146; CA24433.1; -
DR PIR; A04351; HIBPC7.
DR PIR; S42331; S42331.
SQ SEQUENCE 747 AA; 84341 MW; 959C572B7B42C2B8 CRC64;

```

Query Match 3 1%; Score 114; DB 1; Length 747;
 Best Local Similarity 18.8%; Pred. No. 8.3; Indels 218; Gaps 27;
 Matches 114; Conservative 86; Mismatches 190; Indels 218; Gaps 27;

```

Oy 91 GYGATOSSTVDHDXLRDQFVSRTWGEELPIPGLSVKYLRKIVTKDVEGOLKGTDEM 150
Db 26 GYRAATTQAEQPRSSLLD-----TIG-----RAKKGADMTYTKAEORADLAD- 69
Oy 151 ERLKRAQVCEOLA---KRENADENOLCIVEPFS-----NNEYFLIYDVFKDVR 198
Db 70 ---RSNELIRLTPREQREALNNGTLLYODDPAMEALRVKTKGNMAALYVDDDMQKIK 125
Oy 199 M-VFAPSSVGKFGGDTTNMKMPRTIGDFSVFRYA---GADNRAEAKSKDNKPKPYF 254
Db 126 EGVFTREMEBY-----RHSRLQEGAKYVAEQGIDPEDVDYQ----- 165
Oy 255 AAVSQGYKADYATIGFPSTDRYTFSGVEDIEENNPRIEVRIGQ----- 305
Db 166 -----GENGDITTEINILYGAHDFLSQQAOKGAINNS---RVELNVLDDPDMLRPP 216
Oy 306 -----GIWKEANSAD-QATRI-----KYASKYAOSANYWKNKSGNRLGLARDV 348
Db 217 SADFFEKYIDNGLVYGAIPDAQATQLISOAFSDASSBAGADF-----LMR 263
Oy 349 IGRKAERARADMTKNGKSAVYGDV-----SLEKAYK----- 384
Db 264 VGDKR-----VTLNGATTYRELIGEQWALAVTAQRSOPETDAKLNEQYRLKIN 314
Oy 385 -----EGAKA-----NREMTYLSSETLFGTEVVRFAQFANALATNPDA 422
Db 315 SALNEDERTAMEMIQIKAEIDKQYPOEQMTPOREMLISAE-----QYONQMANMTKA 369
Oy 423 HAGILKSLDDKXYKDYLPISLRKRVLPAMLDIV---RRRIPADKLPDIFKNVIDKKFGDT 478
Db 370 QA---KALDDSMK-----SMNK-----LDVIDKQPKRKTINEMVSTDKMPVENTGEF 416
Oy 479 KKYADFVDKSVYPSDGFHMLKSM-----KKFKAKIEKDPAY----- 519
Db 417 K-----HSDMVNYANKRLAEISMDIPDGAOKAMKLYQAOSKDAEFTALGTAVTD 469
Oy 520 ---ELKSVIAAARAIOADMANAYAIKGRLEFAGLREMPGRALPSDANFTMRMSYG 576
Db 470 AGQEWSAVINGKLPERRPPADALRLRINADPQLAL-----YDQAELEPLTMDM 521
Oy 577 SIKGYEPQ 584
Db 522 DKQGIDPQ 529

```

```

RESULT 8
RIRL_CHLMU STANDARD: PRT: 1047 AA.
ID RIRL_CHLMU
AC Q9PL93;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ribonucleoside-diphosphate reductase alpha chain (EC 1.17.4.1)
DE (Ribonucleotide reductase).
GN NRDA OR TC0214.
OS Chlamydia muridarum.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83560;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MOPn / Nigg;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gall S.R., Heideberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwin M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MOPn and Chlamydia
RT pneumoniae AR39."
RL Nucleic Acids Res. 28:1397-1406(2000).
CC -I- FUNCTION: CATALYZES THE BIOSYNTHESIS OF DEOXYRIBONUCLEOTIDES FROM
CC THE CORRESPONDING RIBONUCLEOTIDES, PRECURSORS THAT ARE NECESSARY

```

CC	FORDNA SYNTHESIS (BY SIMILARITY).
CC	-I- CATALYTIC ACTIVITY: 2'-deoxyribonucleoside diphosphate + oxidized
CC	thioedoxin + H(2)O = ribonucleoside diphosphate + reduced
CC	thioedoxin.
CC	-I- PATHWAY: FIRST REACTION IN THE DNA REPLICATION PATHWAY.
CC	-I- SUBUNIT: Tetramer of two alpha and two beta chains
CC	(by similarity).
CC	-I- SIMILARITY: BELONGS TO THE RIBONUCLEOSIDE DIPHOSPHATE REDUCTASE
CC	LARGE CHAIN FAMILY.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/sib-ch).
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; AE002288; AAF39086.1; -
DR	TIGR; TC0214; -----
DR	InterPro; IPR007088; Ribonucleo.red.
DR	Pfam; PF00317; ribonuc_red.1g.1.
DR	PRINTS; PR01183; RIBORDTASEM1.
DR	PROSITE; PS00089; RIBORED.LARGE; 1.
KW	Oxidoreductase; DNA replication; Complete proteome.
FT	ACT_SITE 458 458 BY SIMILARITY.
FT	ACT_SITE 672 672 BY SIMILARITY.
FT	ACT_SITE 687 687 BY SIMILARITY.
FT	SITE 1043 1043 INTERACTS WITH THIOREDOXIN/GLUTAREDOXIN.
FT	SITE 1046 1046 (BY SIMILARITY).
FT	INTERACTS WITH THIOREDOXIN/GLUTAREDOXIN
FT	(BY SIMILARITY).
SO	SEQUENCE 1047 AA; 120035 MW; 9B3FFD9BF817AA CRC64;
QY	Query Match 3.0%; Score 113; DB 1; Length 1047;
QY	Best Local Similarity 18.1%; Pred. No. 15;
MATCHES	111; Conservative 106; Mismatches 196; Indels 200; Gaps 32.
DQ	112 SRTMGCELPIDGLSVKYLRIKYVKNYKGQLGTTDEMERLRKAQEVCDELAKKENADE 171
:	: : : : : : : : : : : : : : : : :
DQ	37 TRRIDDHNP-----EDLENSTRSTHGV--VKEVYQKT-----D 71
QY	172 NOLCIPEPFYSNNYEELIVYDVFKDYRVAFDPSSVGKFGDDTWMMPRHTGDSEVFERY 231
:	: : : : : : : : : : : : : : : : :
DQ	72 GQVVTVVERIQDMVESQLYI-NGLDQV-----ARDYVYVR- 104
QY	232 YAGADNRRAEYSKD-----NRPKYRYFAAASMGCKRDD--YAATIGPFGST 277
DQ	105 ----DDRRAHREKSQSLSVTRRCCTTYHFENPMKISALEKFARATDRIEGTTPDFVREE 160
QY	278 DRYLYTS---GWVEDRIENENNPNRIENVRGIKGQIMKEAMSADQATRLKYASKVAOSANY-- 332
:	: : : : : : : : : : : : : : : : :
DQ	161 VNALTLQTKVALTELEEGCSODS-RIDIETQITDIIVEQLM-----YGHATAAKNTLL 210
QY	333 WKNSIGMNRGLARLDYIGRKRAEEBAFADWIRKNGKSAVYGVLSSLKEYEGAKANRE 392
:	: : : : : : : : : : : : : : : : :
DQ	211 YREA-----RARVDNRVEDQIVEEAPSETFEVLS----KDGSG----- 245
QY	393 MYTILSETLEGGEVVRFAQFNALATNDNAHGILSKSDKKDKDVLPSL-DKQVLPAMLD 451
:	: : : : : : : : : : : : : : : : :
DQ	246 -TYMT-----THSQLLARLARACGRPEPTDAALLT-DMAISNEYSGIKRESEVYLACIM 297
QY	452 IVRRRIIPADKLPD-----IFKNVIDKPKCGDTKKAYADVFDKSVVPYSDKFHA 499
:	: : : : : : : : : : : : : : : : :
DQ	298 AARANI-EKEDBVAFVAABELLDVYYVEALDRS-RQGED----- 334
QY	500 MIKSMDEKKEPAKAIKKDPAVELSKSVIAARAIOADANANAAYAIEKRGLFFAGIREMY- 558
:	: : : : : : : : : : : : : : : : :
DQ	335 -LEQYRDHFHKRYIMEGGSYRLNPFL--KNLFDDALANAMDLSRDLOFSYMGIONLYD 390
QY	559 -----PGRALPDSANFTMRMSYG-SIGKIEYQDGM--YNYITT-----GK 596
:	: : : : : : : : : : : : : : : : :
DQ	391 RFYNHDDRRRLTEPOLTFMKRAMVAGIALK-EQDKTYWATTFYNLLSTERVYPATPTLFS 448

Oy	597	GYLEODKSPDEFAOVENILDFERKNGRAVNOQLIAFIS-----NNDITG-GMSG	649
Db	449	GMRHQLSSCTISLTSTYQDDLVNITKY-----ISDN-----AMLSKWAGSIGDMWIRATG	498
Oy	650	SPVEDKNGRLIGL	662
Db	499	ALIKGTNGKSGCV	511
RESULT 9			
ID	CN16_HAEIN	STANDARD:	PRT: 657 AA.
AC	P44764:		
DT	01-NOV-1995 (Rel. 32, Created)		
DT	01-NOV-1995 (Rel. 32, Last sequence update)		
DT	16-Oct-2001 (Rel. 40, Last annotation update)		
DE	2',3'-cyclic-nucleotide 2'-phosphodiesterase precursor (EC 3.1.4.16).		
GN	CPDB OR H10583.		
OS	Haemophilus influenzae.		
OC	Bacteria: Proteobacteria; gamma subdivision; Pasteurellaceae;		
OX	NCBI_TaxID=727;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-RD / KM20 / ATCC 51907;		
RX	MEDLINE=95350630; PubMed=7542800;		
RA	Flerschmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,		
RA	Kerleyave A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,		
RA	McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,		
RA	Scott J.D., Shirley R., Liu L.-I., Gilek A., Kelley J.M.,		
RA	Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,		
RA	Utterback T.R., Hanna M.C., Nguyen D.T., Saudke D.M., Brandon R.C.,		
RA	Fliege L.D., Fitchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,		
RA	Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,		
RA	Venter J.C.;		
RT	"Whole-genome random sequencing and assembly of Haemophilus influenzae		
RT	Rd.";		
RL	Science 269:496-512(1995).		
CC	-I- FUNCTION: THIS BIFUNCTIONAL ENZYME CATALYZES TWO CONSECUTIVE		
CC	REACTIONS CONVERTING 2',3'-CYCLIC-NUCLEOTIDE TO 3'-NUCLEOTIDE		
CC	AND THEN 3'-NUCLEOTIDE TO NUCLEIC ACID AND PHOSPHATE		
CC	(BY SIMILARITY).		
CC	-I- CATALYTIC ACTIVITY: Nucleoside 2',3'-cyclic phosphate + H(2)O =		
CC	nucleoside 3'-phosphate.		
CC	-I- SUBCELLULAR LOCATION: Periplasmic (By similarity).		
CC	-I- SIMILARITY: BELONGS TO THE 5'-NUCLEOTIDASE FAMILY.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL: U32740: AAC2242.1; -.		
DR	TIGR: H10583:		
DR	InterPro: IPR002224; 5_nucleotidase.		
DR	InterPro: IPR000934; Ser_thr_phosphase.		
DR	Pfam: PF01009; 5_nucleotidase; 1.		
DR	Pfam: PF02872; 5_nucleotidase; 1.		
DR	PROSITE: PS00785; 5_NUCLEOTIDASE_1; 1.		
DR	PROSITE: PS00786; 5_NUCLEOTIDASE_2; 1.		
KW	Hydrolase; Multifunctional enzyme; Periplasmic; Signal;		
KW	Complete proteome.		
FT	SIGNAL	1	POTENTIAL.
FT	CHAIN	27	657
FT	SEQUENCE	657 AA;	72763 MW; 201CABAB15014499 CnC64;
SO	Query Match	3.0%;	Score 112.5; DB 1; Length 657;
	Best Local Similarity	18.2%;	Pred. No. 8.7;

Matches 122; Conservative 82; Mismatches 193; Indels 273; Gaps 36;

QY 187 FLIVDVDFKDVAVFAPSPSSVGKFG-----GDTNNMMWRHTGDF 226
 DB 45 FLTDEPDYID-----AP---TDFGFTTAASTLRQARAAYKKSALVNDGDLIQGNFIADY 96
 QY 227 SVFRYAGADNRPRA-----EYSKDNKPKPYFAVANSOGYK 263
 DB 97 QAAOGYKEGKSNPAIDCLNANMYEVCTIGNHFNGLANTLAD--AIKQAKFPIVNSVYK 154
 QY 264 ADDYANTIFPGSTDRYILTSWGVEDR--IENENRPIEVRGKOG-----IMKKA 311
 DB 155 A-----GTEEPYFPYVYIEKSYVNDG---KTKKLTIGIGFVPPQIMWMDA 200
 QY 312 -MSADQATR--IKYASKYQSANMYKNSIGMNGRLARLDVIGRKRAREPAPADWIRKNC 368
 DB 201 NIGKVEREDYIKTAKYVPE-----MKKG- 226
 QY 369 SAIVYGDVLSL-----EKAYKEGAKAN---REMYTSETLFGTEVY---RPAQFAN 414
 DB 227 ---ADIVVALANTGSPDEPYOEGAENSAPYLADVPHIDAVIFGHSRLEFPNKEFAKSPN 282
 QY 415 ALATN-----PDANAG-----ILKSLDDKXKDYLPISDRKYLPAMLDIVRRIPAD 460
 DB 283 ADIVNGTVAGIPESMAGYANNISVVDGLTEHKGMVITSGAAVLRPIYDIETKALAK 342
 QY 461 KLPDIFKNVIDKKFGKGTADFEVDKSVPYSDKFHMLKSMDEKPAKAIKDPAYE 520
 DB 343 NDEPI--TALLKPVHATRTKRYVSQPIGKA---TDMNYSYL-----ALLQDDPTIQ 387
 QY 521 LSKS-----VIAARAIOADAMAN---AYA-IEKGRLEF-FAGLR 555
 DB 388 IYNQAKAVKAVKAPSIAMAGLPILISAGAPKAGRKNDPTGYEVNKGKLFERNAADL 447
 QY 556 EMVPG-----RALPSPDANTMMSYSGISGYEPQ-----DGANY 589
 DB 448 YXPTPLVYVKAATGEOLKLEKLECSAGMKFOIDTSDKPSGLDMGSEFRYVNDYIDGVNY 507
 QY 590 NYHTT-----GKGVLEKODPKSDEFAVQENIIDLFFTKNY 624
 DB 508 EYDLTFKPARYDGECKLINPESHVVNLTYQKPY---DPKA-EF-----LATNTN 554
 QY 625 GRAYE-----NGOLHIAFELNND-----ITGNSSS--PYDPKNGRLIGLADGW 668
 DB 555 RAYGNKFPGTGKHIVYASPDERSQIADYIKATSEKGSVNPADKNRVPIT--GMD 612
 QY 669 EAMSGDIEFE 678
 DB 613 KL---DVREF 619

RESULT 10
 EST2_CABEL
 ID EST2_CABEL STANDARD; PRT; 556 AA.
 AC 007085; 556
 DT 01-OCT-1994 (rel. 30, Created)
 DT 01-OCT-1994 (rel. 30, Last sequence update)
 DT 01-OCT-1994 (rel. 30, Last annotation update)
 DE Esterase CM06B1 (EC 3.1.1.1).
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Pelodermidae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94033620; PubMed=8219278;
 RA Fedon Y., Cousin X., Toulant J.-P., Thierry-Mieg D., Arpagaus M.;
 RT "cDNA sequence, gene structure, and cholinesterase-like domains of an
 RT esterase from Caenorhabditis elegans mapped to chromosome V.";
 RL DNA Seq. 3:347-356(1993).
 CC -1- CATALYTIC ACTIVITY: A carboxylic ester + H(2)O = an alcohol + a
 CC carboxylic anion.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC. ATTACHED TO THE INNER SIDE OF

CC THE MEMBRANE BY A LIPID ANCHOR.
 CC -1- SIMILARITY: BELONGS TO THE TYPE-B CARBOXYLESTERASE/LIPASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: X66104; CAA46899.1; -.
 CC PIR: S21556; S21556.
 CC HSP: P37967; I0E3.
 CC InterPro: IPR002018; Carboxylesterase-B.
 CC InterPro: IPR000379; Est_lip_thioest_acetate.
 CC Pfam: PF00135; Coesterase; 1.
 CC PROSITE: PS00122; CARBOXYLESTERASE_B_1; 1.
 CC PROSITE: PS00941; CARBOXYLESTERASE_B_2; FALSE_NEG.
 CC DR Hydrolase; Serine esterase; Myristate; Membrane.
 CC FT INIT_MET 0 0
 CC FT LIPID 1 1
 CC FT ACT_SITE 207 207 MYRISTATE (POTENTIAL).
 CC FT ACT_SITE 330 330 BY SIMILARITY.
 CC FT ACT_SITE 445 445 BY SIMILARITY.
 CC FT DISULFD 75 97 BY SIMILARITY.
 CC SEQUENCE 556 AA; 62391 MW; 46BEB5B2428DB99B CRC64;

Query Match 3.0%; Score 112; DB 1; Length 556;
 Best Local Similarity 20.0%; Pred. No. 7.4;
 Matches 77; Conservative 54; Mismatches 126; Indels 128; Gaps 20;

QY 306 GIMKEAMSADQATRIKAYQAQSANMYKNSI---GMNGLARLDVIGRKRAREPAPADW 362
 DB 176 GLM-----DQTLALKWVKHISFEGDPNCVTVFGSAGSGASTDLSLSPHSRDLFGQF 229
 QY 363 IKRNGKS---AVYGDVLSLEKAYKEGAKANREMYTSETLFGTEVYRPAQF----- 412
 DB 230 IPISTAHGDFAI-----RASENQAKIFRE-----FAEFHGFSGR 264
 QY 413 -ANAL-----ATNPDAHA---GILKSLDDKXKDYLPISDRKYLPAMLDIVRRIPADKLP 463
 DB 265 DSSALFKWYQEQSPETLSNVKGYKKSISG-FLTFIPNLDDGDFPKPLDKREAP----- 318
 QY 464 DIFKNVIDKKFKGDTKRY-----ADF-----VEDKSVYPSDKFHAML 501
 DB 319 -----KKQMTGYTEYEGMLASWNPAPFSPADVGLTLPQGIYKDVVSNPDE---I 367
 QY 502 KSMDEKFAKAIKDPAYELSKSVIAAARAIOADAMANAIAIE-----KGRLEFAGL 554
 DB 368 OKIFEKVEYEGVDSDELMMRKLEAL---GDEFFNGVIOAKNAKHNEVEFYTF 423
 QY 555 REMYPCRALPSDANFTMRMSYSGISGYEPQDGAWNYHTT-----GKGVLEKODPKSDE 608
 DB 424 EYVNP-----SFGMMDGMPFKAA---VHCTELRYLLGEGSVYSKFEPTED 467
 QY 609 FAVQENIIDLFFTKNGRYAE-NGQ 632
 DB 468 RKVMETTTTLF--SNPAKYGNNPK 490

RESULT 11
 DP02_YEAST
 ID DP02_YEAST STANDARD; PRT; 1504 AA.
 AC P14284; 1504
 DT 01-JAN-1990 (rel. 13, Created)
 DT 01-JAN-1990 (rel. 13, Last sequence update)
 DT 01-MAR-2002 (rel. 41, Last annotation update)
 DE DNA polymerase zeta catalytic subunit (EC 2.7.7.7).
 GN REV3 OR PS01 OR YPL167C OR P2535.
 OS Saccharomyces cerevisiae (Baker's yeast).
 CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC Saccharomycetales; Saccharomycetaceae; Saccharomyces.


```

OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9000808; PubMed=2676986;
RA Morrison A., Christensen R.B., Alley J., Beck A.K., Bernstine E.G.,
RT "REV3, a Saccharomyces cerevisiae gene whose function is required for
RT induced mutagenesis, is predicted to encode a nonessential DNA
RT polymerase."
RL J. Bacteriol. 171:5659-5667(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=528C / AB972;
RX MEDLINE=97103777; PubMed=8948103;
RA Purnelle B., Coster F., Goffeau A.;
RT "The sequence of 55 kb on the left arm of yeast chromosome XVI
RT identifies a small nuclear RNA, a new putative protein kinase and two
RT new putative regulators."
RL Yeast 12:1483-1492(1996).
CC -1- FUNCTION: REQUIRED FOR INDUCED MUTAGENESIS. IT IS A NONESSENTIAL
CC DNA POLYMERASE. IT MAY FUNCTION IN TRANSLATION SYNTHESIS.
CC TRANSLATION SYNTHESIS IN S.CEREVISIAE MAY USE A SPECIALIZED DNA
CC POLYMERASE THAT IS NOT REQUIRED FOR OTHER DNA REPLICATIVE
CC PROCESSES.
CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + [DNA](N).
CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M29683; AAA34968.1; -
DR EMBL: X96770; CAA65554.1; -
DR EMBL: Z73523; CAA97873.1; -
DR PIR: A33602; A33602.
DR SGD: S0006088; REV3.
DR InterPro: IPR002064; DNA_POL_B.
DR Pfam: PF00136; DNA_POL_B: 1.
DR Pfam: PF03104; DNA_POL_B_exo: 2.
DR PRINTS: PR00106; DNAPOLB.
DR SMART: SM00486; POLBc: 1.
DR PROSITE: PS00116; DNA_POLYMERASE_B: 1.
KW Transferase; DNA-directed DNA polymerase; DNA replication;
KW DNA-binding; DNA repair; Nuclear protein; Zinc-finger.
FT ZN_FING 1398 1417 C4-TYPE (POTENTIAL).
FT ZN_FING 1446 1473 C4-TYPE (POTENTIAL).
SQ SEQUENCE 1504 AA; 172956 MW; 34C2C6B64F734F5 CRC64;

Query Match 3.0%; Score 112; DB 1; Length 1504;
Best local similarity 17.8%; Pred. No. 29;
Matches 146; Conservative 105; Mismatches 255; Indels 316; Gaps 39;

QY 38 NUDRMRELGLTLP-LDSLYSFDKSTINAVYITGGCTGTVSOGGLFTNHHC----- 90
DB 212 NVDRCY---FRSPVLNSILIDDK-----LTINDQLDLDRFCDFKCNV 252
QY 91 -----GVGAIO-----SOSTVDHDLNR-GPVS-----RTM 115
DB 253 LSRDPRVNGNLTIEDILPOFTKNREKLOHRDINHDFLEKLGDISIDIPVKPYSSADM 312
QY 116 GEELPI--PGLSVKYLKKIVKVDKVGOLKGTIDEMERL-RKAQEVCOELAKKENADEN 172
DB 313 INELTMQRELSLKEKPEPKTRHVSQHOMQSGFEAFYKAKQH-----KTSFEDG 365
QY 173 QCLVEFYSNNEFLVYDFKVRNVPAPSSVSGKFGGDTDMWMPRHTGDFSVRRVY 232
DB 366 QIPNFENFIKNOKFSAINPYE-----ALPQ-----LWPR----- 396

```

```

QY 233 AGADNRPAEYSKDNKPKPYVEAAVSMQYKADDYAMTIGFGSTDRYLSM---GYEDR 289
DB 397 -----LPOIEINN-----SMODKKNDD-----QVNASFTYEICGVNDE 431
QY 290 IENENPRIEVRGIKOGIKWKEAMGADATRIKYASKYAQSANV-----WK 334
DB 432 NEGKGSNIKRSYSWLPESIASPKDSYILLDQTKHNTINFNSDMCAMQNMASKRLR 491
QY 335 NSIGMNGRLARDYIGKKRAEERAPADWIRKNGKSANYG-----DYLSSLEKA----- 382
DB 492 SSVSAN---KTSLSRRK--KKVMAGLRYGKRAFYVEGPEPFYODILNKLEDEGFPKI 545
QY 383 -YKE-----GAKANREMYLSTL---FEGTEV-----VRAQAPAN-A 415
DB 546 DYKDPFSSNPVDLENKPKPYAAGKRFELSSHTVSTRIPYQFGEVSYNKKTFPMFSWK 605
QY 416 LATNPDAHAGILKSLDDKYKDYPLDRKVPALMDIYRRIPADKLPDIKKNYIDKKFK 475
DB 606 YALRPPIYDAN-----QKWYKVPYSGMKKTESDISM---HTPHSKF--LYKFPASDVSGK 655
QY 476 GDTYK-----YADFYDKSVVPYSOKFHAKLSMDKEKFA----- 510
DB 656 QKRKSSVHDSLTHTLEIHNATRSKDKIPDPAIDEVSMITICLEETFPDLDIAYEGIM 715
QY 511 ---KAIEK-----DPAYELSKSVYAAARAIOADAMAN-----A 540
DB 716 IVHRASESTFPYKIQHCINEIPVMEFYESEFEDEALDVLDPDILSLGFEIHNFSWG 775
QY 541 YAIKGRKLF-FAGLRE-----MYPG-----RALPSDAN 568
DB 776 YIIRCKQHFDIVRELARVQCQIKTRKLPDWTGYAHSSGIMITGRHINWRLRSQVN 835
QY 569 FTMMSYGSISIGYEPDQAMYNHTTGKYLEKDDPK-----SDEPAVOENIIDLFRTK 622
DB 836 LTO-----YTIESAAFNILHKRLPHSFESFLTNMMAKSTELKTVL 878
QY 623 NYGRYAENGOLHIAFLSNDDITGNSGSPYFDKNGRLIGLAF 664
DB 879 NY--WLSRAQINIOILRKQDIARTI-----EQARLIGIDF 912

RESULT 12
ACON_RICPR STANDARD; PRT; 878 AA.
AC Q92CF4;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Aconitate hydratase (EC 4.2.1.3) (Citrate hydro-lyase) (Aconitase).
GN ACNA OR RP799.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=782;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=MADRID E;
RX MEDLINE=99039499; PubMed=9823893;
RA Andersson S.G.E., Zomori dipour A., Andersson J.O.,
RA Slicheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
RT mitochondria".
RL Nature 396:133-140(1998).
CC -1- CATALYTIC ACTIVITY: Citrate = cis-aconitate + H(2)O.
CC -1- PATHWAY: TRICARBOXYLIC ACID CYCLE.
CC -1- MISCELLANEOUS: ACONITASE HAS AN ACTIVE (4FE-4S) AND AN INACTIVE
CC (3FE-4S) FORMS. THE ACTIVE (4FE-4S) CLUSTER IS PART OF THE
CC CATALYTIC SITE THAT INTERCONVERTS CITRATE, CIS-ACONITASE, AND
CC ISOCITRATE.
CC -1- SIMILARITY: BELONGS TO THE ACONITASE/TPM ISOMERASE FAMILY.
CC -----

```


CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: AJ25273; CA15225.1; -
DR HSSP: P16276; 7ACN.
DR InterPro: IPR001030; Aconitase.
DR InterPro: IPR000573; Aconitase_C.
DR Pfam: PF00330; aconitase_1.
DR Pfam: PF00694; Aconitase_C; 1.
DR PRINTS: PR00415; ACONITASE.
DR PRODOM: PD000511; ACONITASE; 1.
DR PROSITE: PS01244; ACONITASE_2; 1.
DR PROSITE: PS01244; ACONITASE_1; 1.
KW Lyase; Tricarboxylic acid cycle; Iron-sulfur; 4Fe-4S;
KW Complete proteome.
FT METAL 426 IRON (IRON-SULFUR CLUSTER)
FT METAL 426 (BY SIMILARITY).
FT METAL 492 IRON (IRON-SULFUR CLUSTER)
FT METAL 495 (BY SIMILARITY).
FT METAL 495 IRON (IRON-SULFUR CLUSTER)
FT METAL 495 (BY SIMILARITY).
SQ SEQUENCE 878 AA; 97120 MW; 283539115E6805 CRC64;

Query Match 3.0%; Score 111.5; DB 1; Length 878;
Best Local Similarity 19.4%; Pred. No. 15;
Matches 144; Conservative 96; Mismatches 255; Indels 249; Gaps 37;

OY 9 LIGALALGASGAKADKGMLELNQENLDMRELGLPLDLSYSPDKPSIAAVYI 68
DB 242 VIGKLIKLAGAKATATDVLTTETILR-----KKVGFVER 280
OY 69 FGGGCTGIVSDGLFTNHHGCGYAIOSOSIVDHDYLRDGFVSRTMGEELPIGLSVKY 128
DB 281 FCGNGLHNLITSDRATL-SNNSPEYGATCGFPIDQETLK-----YLEVTGREITQ 329
OY 129 LRKIVAVTDKVEGQLGIDEMERLKAQE-----VCOELA--KKNADENOCTIYEPY 181
DB 330 IKLVEYTA--TEONLMYNCEDQETYEVELDISTVYSSLAGPKRQDRVNLNCVANSFO 387
OY 182 SNNEYLIVYDVFKDYRMAPPSVSGKFG-CD-----TDNMWMPRHGTDFSVFRV 231
DB 388 NELPYALAN--KQIDKTYAANONMTAIGNGVVIAITSCI-----NTGNPSVMIG 437
OY 232 YGADNRPAEYSDKRPYKPYEPA-----VSMQGYKADYAMTIGPGSTDRYLS 283
DB 438 AALLAKALEHGLNPKPWVYTSLAPGSKVYTEYKLSG--LDKYLALGF-----NLVG 489
OY 284 WGVEDIEENEN--NPRIEVRGIGKIMKEMASDAQATRIKYASKYASANYMKNISGMNR 341
DB 490 YGCTTGIGSGSLNPELE-----NTINKNR 514
OY 342 GLARLVYIGKRAEERAFADWIKRNGKSAVYGDVLSLEKAYKEGAKANREMYTSETLE 401
DB 515 LVYASVLSGKRNFEGR-----INPLTKA-----SYLSPIL 545
OY 402 GGTVEVRFQAQA--NALATNPDAHAGILKSLDQKRYDPLSLDRKYLPAWLIVRRRIPA 459
DB 546 ----VVAVALSGTLNIDLTFTPIGANIY-----LKDIMP-SQKEIDVANSINSMTI 594
OY 460 DKLPDLFKNVIDKKKGDTKKYAD-----FVDFKSVV--PYSDKFAH--MLKSM 504
DB 595 EKADVDENG-----TKEMRDLOITGTNTNMDKNSYINNPPYFDNIGSEHSIKDI 645
OY 505 DKKEK-----FAKATEKD--PAVELSKSVTAARAIOADAMA--NMYALEKGR-----RL 549
DB 646 KSKRIILAIAGDSITTDHISFAGSISKNSPAKYLKHHLEPDLFNSIGSRGHEVMNRG 705
OY 550 FFAGLR---EMYPGRALPDPANFTM-----RMSYGSIKGYEPD-----G 586

DB 706 TFIANIRKKNEMCNG-----VEGGFTINQLSGVQOTIVDFADWDYKAHDIPLVIFAGKEVGS 761
OY 587 AWNYHTTKGVLEKODRPSDEFRAVENIIDLFTRTNMYGRANQGLHIAFLSNNDITG 646
DB 762 SSRDWMAGKGGLL-----GIKAVIAESPERIHSNLSVGMILPLTF-----TGK 805
OY 647 NS-----GSPVFDKNGRLIGIA 663
DB 806 NTRLNKLKDGSEIID-----LIGLS 825

RESULT 13
YKLE.CAEEL STANDARD; PRT; 922 AA.
ID YKLE.CAEEL
AC P42173;
DT 01-NOV-1995 (rel. 32, Created)
DT 01-NOV-1997 (rel. 35, Last sequence update)
DT 30-MAY-2000 (rel. 39, Last annotation update)
DE Hypothetical 103.8 kDa protein C03C10.6 in chromosome III.
GN C03C10.6.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Felodierinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2.
RA Gardner A., Berts M.
RL Submitted (Oct-1994) to the EMBL/GenBank/DBJ databases.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: Z38112; CAA86235.1; -
DR EMBL: Z35637; CAA84690.1; -
DR EMBL: Z38112; CAA84690.1; JOINED.
DR WormBep: C03C10.6; CE02959.
KW Hypothetical protein.
SQ SEQUENCE 922 AA; 103818 MW; DEB088AB97FC87D8 CRC64;

Query Match 3.0%; Score 111.5; DB 1; Length 922;
Best Local Similarity 21.1%; Pred. No. 16;
Matches 130; Conservative 77; Mismatches 189; Indels 219; Gaps 33;
OY 55 YSPDKPSIANAVYIFGGGCTGIVSD-QGLFTNHHGCGYAIOSOSIVDHDYLRDGFVSR 113
DB 151 YRFDQLPVSEYIV-----RLGIVYKTLQNOV--HCCSDAADSADIDTWDEDDVAN 203
OY 114 TWGEELPIP-GLSKYKLRKIVKTVTKRVEGOL-----KGTDEMERLKAQ 157
DB 204 A---ELIRPRESLV---TTIGETPAAEAGVNHGAVCMQPTIPEEGGLAPRELIRKRS 257
OY 158 EYCOELAKKENA-----DEN--OLCTIV-----EPYSNNE 185
DB 258 TAVQEVRRKKSALAEVKKSCCTLRGCHNEAIEFDEETISLCLIALEFELHREPSMAADP 317
OY 186 YPLIYVDYFKDYRMAPPSVSGKFGGDDNNMMPRHGTDFSV----- 228
DB 318 LFKILYTV--TRLIDTP-----MYPHSTEMFVPANSRSVAKQMLRVSLH 360
OY 229 -----FRVYAGADNRPAEY-----SKDKRPYKPYFAVSMQGYKADYAMTIGF 273
DB 361 HLTSAICQLPFTKIPRDAFMSVALSLADPELSPYFIQIIMEDLEE-----SW 413
OY 274 PGSTD---RYLSWGVEDRIENENNPRIEVRGIGKQIKW--EAMGADA--TRIKVAS 324

```

Db 414 PGSVKLMKMLAFYIEPTDMYNPNPKDVLGHLEFFFRKYHSAISADNGITPTBRAE1-- 471
QY 325 KYQSAHYMKNSIGMRGLARLDVIGKRAE---ERAFAMWIKNGKSANYGVLSLEK 381
Db 472 -----ENVITVMTHEVKYQTFSSSKSPVTLVEFAFMLESLSLA---DV--SLES 517
QY 382 -----AYKGAKANREMTYLTSETLFGTEVVRFAOFAANL-ATNPDAHGIKLSDDX 434
Db 518 LIGVCAKCNALRLRENDKOCITRALV--TELMOAIFKVKVLSHNSVYTIANMT--LQDAG 573
QY 435 KDV-LPSLDKRVLPAMLDIVRRRIPADKLPDIFKNVIDKKDKGDTKKYADFVDSVVPY 493
Db 574 EDIEVPLDDQFNTFASEAIR-----PELF-----EVLDEFIDLHYI-- 610
QY 494 SDFHMLKSKMD-----KEKFAKATKEDPAVELSK-----S 524
Db 611 -----AKLKESNSDALGDLKVKLAELAI---AVEMSRNARDCRTVIRIFWLMSPPS 661
QY 525 VIAAARAIOADAMAN 539
Db 662 VTQAASAFADSVTN 676

RESULT 14
MOK1_SCHPO STANDARD; PRT; 2410 AA.
ID MOK1_SCHPO
AC Q9USK8; Q9UR5; Q9URK2;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cell wall alpha-1,3-glucan synthase moki (EC 2.4.1.183).
DN MOK1 OR AGS1 OR SPC338.01C OR SPC1717.01.
OS Schizosaccharomyces pombe (fission yeast).
OC Baktayota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN (1)
RP SEQUENCE FROM N.A.
RA Katsuyama S., Dai H., Arellano M., Perez P., Toda T.;
RT "Fission yeast alpha-glucan synthase Moki localizes closely with actin
RT and play a role essential for cell morphogenesis and protein kinase C
RT function."
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RN (2)
RP SEQUENCE OF 1-1210 FROM N.A.
RC STRAIN=972;
RA Lyne M., Wood V., Rajandream M.A., Barrell B.G., Murphy L., Harris D.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RN (3)
RP SEQUENCE OF 1157-1465 FROM N.A.
RC STRAIN=972;
RA Wood V., Rajandream M.A., Barrell B.G., Saunders D., Harris D.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: UDP-glucose + [alpha-D-glucosyl-(1,3)](N) =
CC UDP + [alpha-D-glucosyl-(1,3)](N+1).
CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 1.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB019183; BAA34054.1; -
CC EMBL; AL023781; CA19332.1; -
CC EMBL; AL049472; CAB39330.1; -
CC InterPro: IPR000461; Alpha_amyase.
CC InterPro: IPR001296; Glycosyltransf_1.
CC Pfam: PF00128; alpha-amyase; 1.
CC Pfam: PF00534; Glycosyltransf_1; 1.

```

```

KW Cell wall; Transferase; Glycosyltransferase.
FT CONFLICT 256 256 F -> Y (IN REF. 2).
SQ SEQUENCE 2410 AA; 272120 MW; 8BD944BAE3A3A3C5 CRC64;
Query Match 3.0%; Score 111.5; DB 1; Length 2410;
Best Local Similarity 17.5%; Pred. No. 58;
Matches 131; Conservative 132; Mismatches 232; Indels 253; Gaps 37;
QY 88 HHCYCAIQSQSTVDHDIADGCVSRKMGELIPGISVY-----LRTYK 134
Db 1355 HOKFGAV-----GVSNKYGRSWARPIFWGLKTKG 1387
QY 135 V-----TDK-VEGOLKGTID--EMERLR-KAOEVCOLAKKENDEMOQIVPEFYS 182
Db 1388 LPPNDPTDDEIVDQKVAITLIDDPDEKSKVEKRIAOQWAGLEVENKIDLVFPGRWIS 1447
QY 183 NNEFFLYVDVDFD-----VRNVFAFP--SVYKFGGDTDNMMWPRH--TGDFSVER-- 230
Db 1448 SQKGIDLIADIASLSEYKVLICVGPILIDYGFKAEXLDVLOKKYPTRVFSQPRFTQ 1507
QY 231 ----VYAGADNRAEYSKDKPKPYFAVSMQ-GYKADDIYAMTIGFPGSTBRLYLSWG 285
Db 1508 LPPYIFSGADFALI-----PSRDEPFGLVAVEFGKRG--ALGIGARVGLGQMPGWW 1557
QY 286 VEDRIENENNPRIEVRGIGKIGKEAMASADQATRIKYSKVAO----- 328
Db 1558 YS--VESSATPHL-LKQFOEAC-QQALSSQRTARARAKSAKORFPVSGMKALEALTD 1613
QY 329 ----SANYWKNSTG-----MNRGLARLDVI-----GRKRAEERAFADWIRKNG 367
Db 1614 GCATCSQKQYVNRSSRSFSLHESFSRSSEVLPSTSDTNLAKRAEE--AEMIMET 1670
QY 368 KSAVYGDVLSLEKAYKEGAK-----ANREMTYLTSETLFGTEVVRFA 410
Db 1671 PPTAEANTGAKLRSLSLGSRGPGATTEDDASDGLDTIOEESMTAGDSTSGSDISRY- 1729
QY 411 OFANALATNPDAHAGILSKSDDDYKDYLPGLDRKVLVLPAMDIVRRRIPADKLPDIFKNVI 470
Db 1730 ----RAERLNPDSSHSPSEYSTDG----- 1749
QY 471 DKRFKGTFRKYADFV-FDKSVPEYSDKFHMLKSMDE---KFAKATKEDPAVELSKS 524
Db 1750 DYEFDPQRSYYDDLFDDDTTIRNAPSFPRQMSFDEHVAVGATFSQDDLSDP----- 1802
QY 525 VIAAARAIOADAM-----ANAYATEKGRLPFAGIRENTPGALISDANFTWRMS 574
Db 1803 ----ARVSDSDSVSPPLPPFVAGSNPNARNNNNPFYGNLH-----TESLSLA-- 1847
QY 575 YGSIKGYEPQDGAWYNYHTT-----GKV-----LEKODPKS--DEFAVOENILDLFT 621
Db 1848 -SAMSCKEKRFSLTRVEEFTDEDGALRSFSEKLOKLANKSKDOLCIEQYIAMSERS 1906
QY 1907 FFEHERRAIKIGLOKPKNLHVELSHSGTESESLNSGQTSYDD---IIMTDESNTYOL 1963
QY 672 SGDIPEPDDORTLSVDIRVLEMDKW 699
Db 1964 -GDQDF-----KTIHGLKMFLEKFIYDW 1985

RESULT 15
SYL_MENJA STANDARD; PRT; 1039 AA.
AC G58357;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Isolation of tRNA synthetase (EC 6.1.1.5) (isoleucine--tRNA ligase)
DE (IleRS).
GN ILES OR MJ0947.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;

```

```

OC Methanococcus.
OX NCBI_taxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA But C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerevage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weisslock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii."
RT Science 273:1058-1073(1996).
CC -I- CATALYTIC ACTIVITY: ATP + L-isoleucine + tRNA(Ile) = AMP +
CC diphosphate + L-isoleucyl-tRNA(Ile).
CC -I- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -I- SUBCELLULAR LOCATION: Cytoplasmic.
CC -I- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb.ch).
CC -----
DR EMBL; U67538; AAB98949.1; -.
DR HSSP; P56690; 1ILE.
DR TIGR; MJ0947; -.
DR InterPro; IPR002300; tRNA-synt_1a.
DR InterPro; IPR001412; tRNA-synt_1.
DR InterPro; IPR002301; tRNA-synt_1le.
DR Pfam; PF00133; tRNA-synt_1; 1.
DR PRINTS; PR00984; TRNASYNT1LE.
DR PROSITE; PS00178; AA-TRNA-LIGASE_1; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Metal-binding; Zinc; Complete proteome.
FT SITE 45 55 "HIGH" REGION.
FT SITE 600 604 "KMSKS" REGION.
FT BINDING 603 603 ATP (BY SIMILARITY).
SQ SEQUENCE 1039 AA; 122233 MW; 7163C46B4AE08AC0 CRC64;

Query Match 3.0%; Score 110.5; DB 1; Length 1039;
Best Local Similarity 19.3%; Pred. No. 22;
Matches 113; Conservative 73; Mismatches 202; Indels 199; Gaps 28;

OY 153 LKAOEV-COELAKKENDENQICVEPYSNNEFLY--YDVFQDVRMVFAPSSVG 208
DB 1 MKKVEPVNERELDKK-----IKKFEWENDIYOKYKKKNRKKKEFYVDGPPYCSG 50
OY 209 KFGDPTDMMWMPRHGTGDSVFRVYAGADNRPAEYSKDNKPYKPYFAAVSMGYKADYV 268
DB 51 AI-----HLG-----TAMNKIIKDYLRFRKRMGGYNVLDKA 81
OY 269 MTIGFPGSDRLRTLSMG-----VEDRIENE---NNPRIEVR-GIKO----- 305
DB 82 -----GMDHGLPIEVKVENEGIKKKKEIETKIGVQKFEIECKEPEALKHK 127
OY 306 -----GIMKEMSADQATRIKAYSKYQASANYMKSISGMNGLARLD--VIGRRK 353
DB 128 EIMKQFKNLGLVGLDMENAYMPTKEY-----MEIGWTLAKVAHEKGLTRDLRVYYWCP 182
OY 354 AEERAFADWIRKNGKSAVYGDVLSLEKAYKGAKANREMTYLSLTFEGTEVVRFAQFA 413
DB 183 RCEALAEHEVGEYKEYVD-----PSYVYKFRLANEENTYI--VIMTTTPWTLVANL- 233
OY 414 NALATNPD-AHAGIILKSLDDKDKYLPISLDRKVLPAMLDIYRRIRPADKLPIDIFKNVIDK 472

```

```

DB 234 -AVTVHPDYDAVVEVEFDDKKEWI--IAEKLVEVIN-----KAKKFNINKYKTIK 284
OY 473 KFKGDTKKYADFVFDKSVVPSYSDKPFAMLSMDKKEKFAKAEKDPVELSKSVIAARAI 532
DB 285 KVKGELEGIKYI-----HPLLENERQKKEFAELENAITVLGEHV----- 325
OY 533 QADAMANAIAIEKGRRLFFAGIRENYPGRALPSDANFTMRKSYGSIKGYEPD----- 585
DB 326 -----TLEGG-----TGLVHTAPGH---GEEDFEVGKKY-NLPYISPIDDEGKYVE 367
OY 586 GAWNYHTTGKGVLEKODPKSDEFAVENIDLFRTKNVGRYA----- 628
DB 368 GKW-----KGVFKD-----ADAEIETTLKNGLLVYAGIKIKHSYPHCWCKTPL 412
OY 629 -----ENGQIHIAFLSNNDITGNSGS--PYFDKNGRLIGLAFDGNW 668
DB 413 LFRATEQWFLSKIKDNIIEHAKTVQMPHNVETRIYNGVVFQDW 459

```

Search completed: October 17, 2002, 22:37.46
Job time : 32 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 17, 2002, 21:41:56 ; Search time 68 Seconds
(without alignments)
1811.359 Million cell updates/sec

Title: US-10-008-355-2
Perfect score: 3719
Sequence: 1 M0MKRLSILGALLIGASG.....LFMIDKGCPRRIQELKLI 712

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match Length	DB ID	Description
1	907.5	24.4	716	16	Q9PC94
2	146.5	3.9	1368	16	Q99ZWW2
3	133.5	3.6	1104	2	Q06812
4	126	3.4	953	2	Q48502
5	125	3.4	1440	5	Q9TYU6
6	124	3.3	1808	3	Q9LQJ8
7	123.5	3.3	713	2	085051
8	123	3.3	1289	16	Q9X087
9	122.5	3.3	742	2	Q9ZB47
10	122.5	3.3	1571	12	Q92B35
11	122.5	3.3	2204	12	Q99F11
12	122	3.3	721	16	Q92BR7
13	122	3.3	743	16	Q9PPY9
14	122	3.3	883	16	Q97S45
15	121	3.3	1208	10	Q9LUP5
16	120	3.2	666	17	Q27798

17	118	3.2	698	16	Q9KU27
18	118	3.2	882	16	Q99XS1
19	117.5	3.2	2204	12	Q99F17
20	117	3.1	912	2	006915
21	116.5	3.1	1273	12	Q9QSK2
22	116	3.1	2080	3	Q92215
23	115.5	3.1	1154	16	Q92K97
24	114.5	3.1	561	16	Q9KB24
25	114.5	3.1	1198	5	Q9Y059
26	114.5	3.1	1306	2	Q47766
27	114.5	3.1	1530	16	Q9KC46
28	114.5	3.1	1758	5	Q22830
29	114	3.1	711	2	Q51284
30	114	3.1	901	2	Q9XCV8
31	113.5	3.1	765	16	Q9TFP1
32	113.5	3.1	1171	5	Q00936
33	113	3.0	632	17	Q97723
34	113	3.0	759	16	Q92A91
35	112.5	3.0	767	16	Q9JYF6
36	112.5	3.0	957	17	Q28565
37	112	3.0	554	5	Q16351
38	112	3.0	804	10	Q9LXB6
39	111.5	3.0	1164	16	Q9ZJN3
40	111.5	3.0	3223	2	Q9RM48
41	111.5	3.0	7659	5	Q9N4M4
42	111	3.0	691	5	Q9GSD1
43	111	3.0	972	5	Q9U5P5
44	111	3.0	1092	2	Q54099
45	111	3.0	1713	9	Q94MA1

ALIGNMENTS

RESULT 1	
Q9PC94	PRELIMINARY; PRT; 716 AA.
ID	Q9PC94
AC	Q9PC94
DC	01-OCT-2000 (TREMBLrel. 15, Created)
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE	HYPOTHETICAL PROTEIN XF1887.
GN	XF1887.
OS	Xylella fastidiosa.
OC	Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC	Xylella.
OX	NCBI_TaxID=2371;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN-9A5C;
RX	MEDLINE-20365717; PubMed-10910347;
RA	Simpson A.U.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA	Alvarenga R., Alves L.M.C., Araya J.E., Bala G.S., Baptista C.S.,
RA	Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Brites M.S.,
RA	Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carier H.,
RA	Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Carne C.M.,
RA	Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA	Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferto J.A.,
RA	Fraga J.S., Franca S.C., Franco M.C., Fritme M., Furlan L.R.,
RA	Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA	Ho P.L., Hohelsel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA	Krieger J.E., Kuramae E.B., Laigret F., Lambais M.R., Leite L.C.C.,
RA	Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA	Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Martino C.L.,
RA	Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA	Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA	Moan D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA	Nhani A.Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA	de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA	Peixoto B.R., Pereira G.A.G., Pereira H.A.Jr., Pesquero J.B.,
RA	Quaglio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA	de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.F.,
RA	da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,

RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshaho M.H.,
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
 RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.,
 RT "The genome sequence of the plant pathogen *Xylella fastidiosa*,"
 RL Nature 406:151-159(2000).
 DR EMBL: AE004008; AAF84693.1; -
 KM Hypothetical protein; Complete proteome.
 SQ SEQUENCE 716 AA; 79375 MW; EAF086E2315BDFC CRC64;

Query Match 24.4%; Score 907.5; DB 16; Length 716;
 Best Local Similarity 31.4%; Pred. No. 9.4e-52;
 Matches 239; Conservative 135; Mismatches 291; Indels 95; Gaps 21;

OY 1 MWMKSLTILGALLGASGVAKADKGMWLNELNOENDRRELGFPLDLSLSPDKP 60
 DB 1 MRFNLISLVLTITVDS--THAGEGMWVPOOL-PELAGPIKQGLQSLPQLSLNLGCD 57
 OY 61 SIANAIVIFGGCGTITVSDQGLIFTHHCGYGAIOSQSTVDHDLRDGFVSRTGSEELP 120
 DB 58 PGAVAVSL--GNCPTASVSPGLVITNHHCAVGAIQLNSTPKNLKIEGFNLQTQADEVS 115
 OY 121 T-PGLSVTKRLIVYTDKVEGOLGITDEMERLKAQEV--QELAKKENDENQCLIV 177
 DB 116 AGPNARIVYLEQITDVTQAKAALAAAGNDPFRRTTALETFSKOEIAKCEE-EGGYRCOF 174
 OY 178 EEPYSNNEFFLVYDVPKDVRFVAPSPSSVGFEGGDTDMMMRHTGDFSVRYVAGADN 237
 DB 175 FEFAGNTYRVKNEIKIDVRLVAPQGSVGFEGGDVDMMPRHTGDSFYRAYVAGKD 234
 OY 238 RPAEYSKDNPKPYKPYFAAVNSOGYKADYATIGFPGSTDRYLTWSGEDRIENENNR 297
 DB 235 KPASFSEKNIYPRPKHMLKFSQDPLGDDGFVAVAGYPRGTRNYALV-----AEPENTAH 288
 OY 298 IEVRIKIQ-----GIMKEAMGADATRIKYASKYQAQSNYKNKSTIGMNGRLARDVIGR 351
 DB 289 WTYPIVIGHFKNLIALIEAASKONPDIOKYASTLAGLNNTSKNDGOLDGRRRNALIGQ 348
 OY 352 KRAEERAFADWIRKNG---KSAVGYDLSLEKAYKEGAKANREMTYLSFTLFGCTEVY 407
 DB 349 KOSEETAVLAWIKOQGINHEALAHQTLVLDLEQY---KANQDRPIVQGF-FNGSGVI 403
 OY 408 RPAQANALA---TNPDAH-----AGILKSIDDKYKYDLSLDRKV----- 445
 DB 404 GYAVNMLYRLAERTSDAQREAGYQERDLPTEGNIKQWE---RRYLDPMDRQMOQYMLT 460
 OY 446 ----PAMLDIYRRRIYA-----DKLPDIEKNVIDKKFGDTKKYADVFPKSVVPSYD 495
 DB 461 EYNKLP---VKQYVALDVMVLGDIPTLRLGDTKLSSSEER----- 500
 OY 496 KEHAMLKSMDEKKEPAKALEKDPAVELSKSVIAARAIOADAMANAAYALEKGRLEFAGIR 555
 DB 501 ----LKMFNADRAAFESSQDPAIRYAVAIMPALLEIEONKIRFGELLKAPITLQALA 555
 OY 556 EYRP--GRALPDANFTMRMSYSGIKYEPQDGAANYNHTTGKYLEKODPKSDPAVOE 613
 DB 556 DYNKSHGKVVYDPAANSIRITFGHWKYSPKDGVEXTEPTTLQGYMAK--NTEVEPEDSBK 614
 OY 614 NLLDFPFRKNGRYAEN--GQHIAPFLSNDDITGNSGSPYEDKNGRLITGLAFQDNWEM 671
 DB 615 SLINAIKAKSVIANLDQRTGIVPVNLSLDDITGNSGSEYLDANIGKLGALFADGMSWV 674
 OY 672 SGDIEEPDLQRTISVDIRYVLEMDIKWQCPRLLOELKL 711
 DB 675 SSNMVFDPMVMTRTIADVSKYQVMTEVAAPAPHLKELNL 714

S1D26

RESULT 2
 O992W2 PRELIMINARY; PRT; 1368 AA.
 AC O992W2;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE HYPOTHEICAL PROTEIN SPY1046.
 GN SPY1046.
 OS Streptococcus pyogenes.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1314;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SE370 / ATCC 700294 / SEROTYPE M1;
 RX MEDLINE-21192684; PubMed-11296296;
 RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic G., Lyon K.,
 RA Pirmeaux C., Sezate S., Suvorov A.N., Kanton S., Lai H.S., Lin S.P.,
 RA Qian Y., Jia H.G., Najaf F.Z., Ren Q., Zhu H., Song L., White J.,
 RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
 RT "Complete genome sequence of an M1 strain of *Streptococcus pyogenes*,"
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
 DR EMBL: AE006549; AAK33936.1; -
 KM Hypothetical protein; Complete proteome.
 SQ SEQUENCE 1368 AA; 158440 MW; 07D04F0B5965762F CRC64;

Query Match 3.9%; Score 146.5; DB 16; Length 1368;
 Best Local Similarity 19.5%; Pred. No. 0.8;
 Matches 167; Conservative 105; Mismatches 228; Indels 357; Gaps 43;

OY 35 NOENIDRRRELGFPLDLSLVSFDRPSIANAVIFGGCGTITVSDQGL--IFTNHHCGY 92
 DB 609 NEENIDLEDIVLTITLFE---DREMIERLKYTAH-----LPDKYMKQLKRRRYIGW 659
 OY 93 GAI-----OSQSTVDHDLVLR-DGVSR-----TWGEELPIGSL----- 125
 DB 660 GRLSRKLINGIRDKSGTII-LDFLKSDFGANRNMQIHDLSLFKEKDIOKAQVSGOGD 718
 OY 126 -----VKYLRKIVKYTD--KVEGO-----LKGITDEM 150
 DB 719 SLHEIHANLAGSPAIAKKGILQTVKVDLVKMGHKNRENIYEMARENQTTQKQKNSR 778
 OY 151 ERLRKAQVCELA---KENADEN-QLCIVEPFSNNEFLIVY-----DVEKDVRYVF 201
 DB 779 ERMKRIEIGIKELSGQILKEHVENTQL-----QNEKLYLYLQNGRQMDYVQDELDI 830
 OY 202 APPSSVGFEGGDTDMMPRHTGDSFVSRYVAGADNRPAAEYSKDNPKPYKPYFAAVNSOG 261
 DB 831 -----NRLSDYDVHD-----IVQOS 845
 OY 262 YKADYAMTIGFPGSTDRYLTWSGEDRIENENNRPIEVRGIGKQIMKEAMSADQATRIK 321
 DB 846 FLKDD-----SIDKVLTRSDKNRKGSDNVSEEVYKKMKYMKQLNAKLITQK 896
 OY 322 YASKYAQSANWYKNSIGNRGLARDLVIG---RKRAEERAFADWT-----RKNGK 368
 DB 897 F-----DNILKABRG---GLSELDKAGFIKRLQVETROTIKHVAQLILDSMNTKYDEN 946
 OY 369 -----SAVYGDVL-----SLEK 381
 DB 947 DKLIREVAVITLKSIVSDPRKDFQFYKREINNYNHAADAYLNAVVTALIKKYPKLES 1006
 OY 382 AYKEG-----AKANREMTYLSFTLFGCTEVVFAQANALATNPDAHAGILKS 429
 DB 1007 EYVYGDYVYVYRKMAIASQEDIGKATAKYFFYSINIMFKEITELANGELIRKRLIETN 1066
 OY 430 -----LDCKYKDYLPSSLDKRYLP-AMLDIYRR-----IP--ADKLPDIFK 467
 DB 1067 GETGEIVMDKGRDF--ATVRYKVLSPQVNIYKTEVGTGFSKESILPKRNSDKL----- 1119
 OY 468 NVIDKKFKGDTKKYADVFQDSVSVYSDKFA-----MKLS-----MKKEFA 510
 DB 1120 --IARRKDWDPKRYGG--FDSEPTAVYSVLVYAVYKVGKSKKLYSKELLGITIMRSSFE 1175
 OY 511 -----KALEKDPAVELSKSVIAARAIOADAMANAAYALEKGRLEFAGLRMY 558
 DB 1176 KNPIDFLAKKYEYKVDLIIKLPR-----YSLFELENGKRRKLASAGELQ 1221

```

Db 699 NNYQYDNR-----YDFMWDQVEGNEILLAHTKILNER-EGYSKYFAR--GERTILVGS 750
QY 496 KFHAAKLSMDEKEKFAKALEKDPAYELSKSV--IAAARAQADAMANA-VAIEKGRLEF 551
Db 751 KDOFLFLFRDYQODKKVYVGLANVAEE-SKAVTLTVDSADAVTVDAVSGREYATAGK---- 805
QY 552 AGLREMYGRALPSPANTMRKSGSYGK-----YEPQGANYN 590
Db 806 --VNLTLTGKADPGCVLLT--VEGNTITGAKNGEVEVVELVPENNIRIHKREDNRYKN 861
QY 591 YHTTGKGVLEKODKPSDEFAVOENTLDFRTKNGRYAE----NGOLHIAFLSNDDIGG 646
Db 862 Y---GATLMNDVYASPSANWPGATGFE--KIDSIGATIDVPLADCAKKTIGLY-MDTIAG 915
QY 647 NSGSPVDPKNGRLIGLAFEDGNWEMASGD--IEFEF-DLQRTISYDIRVLEMD----- 697
Db 916 DAGKDGQGGKGTTSIPQANETWIKGSGKYVTEYEPVLDLPAN-TVRIHTREAVDYDRCGI 974
QY 698 -KMG 700
Db 975 WNWG 978

RESULT 4
Q48502 ID Q48502 PRELIMINARY; PRT; 953 AA.
AC 048502;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)
DE ALPHA-AMYLASE.
GN AMYA.
OS Lactobacillus amylovorus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_Taxid=1604;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CIP 102989;
RA Giraud E., Cuny G.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-155 FROM N.A.
RC STRAIN-NRRLB4540;
RX MEDLINE=95077354; Pubmed=7986030;
RA Fitzsimons A., Hols P., Jore J., Leer R.J., O'Connell M., Delcour J.;
RT "Development of an amylolytic Lactobacillus plantarum silage strain
RT expressing the Lactobacillus amylovorus alpha-amylase gene.";
RL Appl. Environ. Microbiol. 60:3529-3535(1994).
DR EMBL: U62096; AAC45781.1; -.
DR EMBL: X80271; CAA56557.1; -.
DR HSP: P00691; IBAG.
DR InterPro: IPR000461; Alpha-amylase.
DR Pfam: PF00128; alpha-amylase; 1.
DR Pfam: PF02806; alpha-amylase_C; 1.
SQ SEQUENCE 953 AA; 104990 MW; F6ABDEDE5619C5F3 CRC64;

Query Match 3.4%; Score 126; DB 2; Length 953;
Best Local Similarity 19.0%; Pred. No. 10;
Matches 115; Conservative 83; Mismatches 246; Indels 162; Gaps 27;

QY 207 VGKRGCDT-DMMWMPRHNGDPSVFRVYAGADNRPAEYSKDKKPYKPYVEAAVSMOGYKAD 265
Db 95 VGNSSGSKLNNWMLYQPTKYSIGNYILGTEAEFKSMCAAKEVIRIIVDATLMDTSD 154
QY 266 DYAMTIGPGSDT-----RYLTSMGVEDRIE-----NENNPRIEV-----RGI 303
Db 155 YSAISDEIKSISDWTGHWGTQISNMSDREDYQNSLGFYDWNWQISQVYFLKNHLERLI 214
QY 304 KQGIWKEMASDQATRIYKASKY--AGSANWKKSGICNNGRLARLDVIGKRAERARAD 361
Db 215 SDG--ASGFRDADTHIELPSQYDGSYSNSWPN----- 246

```


OY	57	FDKPSIANAAVYIFGGCGTGLTVDGGLIFLNHHCIGGALQOSYSTDHDVLRGCPVSRMG	116
Db	703	WDN-----	731
OY	117	EELPPIPGLS-----VKYLRKIVATVD-----	158
Db	732	LEFLQVVGSTQKADDMKMKCKQVRAHLSLPEEKKLLMDLTAIGRLSMLDILIRKLQIM	791
OY	159	VCQELAKKENADENQLCIVPEFYSNNNEFLIYDVYKDYRAMVPAPSSVGKEGGDTNMM	218
Db	792	VSSRLRDE-----IEEKYANLTH-----	831
OY	219	WPRHTGDSVSRVYAGADNRA---EYSKNNKRYKKVYPAASVMQCYKADDAYMITGPFG	275
Db	832	-----SLDFEPRIRHPEFLLSNRAVDVETWLTLEYCYAADHRAAKLAIFPG	876
OY	276	STDRLT--TSMGVEDHLENNENNPRIEVR-CIGKQIWKEMS-----	313
Db	877	SVYQGVFFPFSNMSADRWYWTTEQRAKLLAKRAIDE---KEKLSKECEKTAKLNLTEQV	933
OY	314	-----ADQATRIKASYKSAOASANYMKNSIGMNGLARLDVIGCKRAE-ERAPADWIRK--	365
Db	934	MHYVAAKHGRKVRKSKSDKHLLA-IQNSSSSSSG-----KKRKTQITKTTGEGVRSII	984
OY	366	-NGSASAYYGDV-----LSSLEKAYKAGKAAKANEYTLSETLFGTEVYVRAPQANA	415
Db	985	VDGKVLNMSDAIDASENKEFLNSLEEHOENIOLENSEIRDLTEDEGOCSSIIIN--QVASS	1042
OY	416	LATN-PRDAHAGILTKSLDDK---YKQDLPGLDKRY-----LPALMDIVRRIRPAD	460
Db	1043	KTTSTPQSRSEWDEADRKILLTSQVYHRAALGAKFLPGVMMAVSPELPAPPLACKRRVQIL	1102
OY	461	KLPDIFKKNVLDKKFKGDTKKYKADFVEDK-SVPEYSDKFHAKLKM-----	504
Db	1103	MKNKFKRAIISLCNLISERYARHLETKQCLPESNKSHTVLYLSIPAIGTDSGVSQEG	1162
OY	505	-----DKKKFKALKK-----DPAVLTLSR-SVIAAARAIQADAMANAIALEKGRKLF	551
Db	1163	KDICEKMDKWDENEKSIQAFNDVLETKKMAKLVAPKRTSSREWSNDDIIEGS---	1218
OY	552	AGLEMYEGRALPDADFTMMSYGSIGKYPEQDQAMVYHTTGKGVLEKODPKSDEFAY	611
Db	1219	-----EMVPAIHSD---IQNVSDQYKQIDSRSSG-HYRLHQVRLRDEKD---NDSIQY	1262
OY	612	QE-----NLL-DLFRTKNYGRYAENGOLHIAELSNND--	642
Db	1268	RKSLAVSTAELTKLVLSMPTAPAGMNLLEDTLR-----RXSER-DLFTAVSYLRDKKF	1322
OY	643	ITGNSGSP-YEDKN-----GRLLGALFQGNWAMSGDIEFEPPDQ	682
Db	1322	LVGSGGQPFVLSQNFILHSISKSPFYVNTGTTRAKFSSWLFIEHERDLMAAGVTLTSDIQ	1380
RESULT 7			
085051			
ID	085051	PRELIMINARY;	713 AA.
AC	085051;		
DT	01-NOV-1998 (TREMBLrel. 08, Created)		
DT	01-NOV-1998 (TREMBLrel. 08, Last sequence update)		
DT	01-JUN-2001 (TREMBLrel. 17, last annotation update)		
DE	TRANSPERRIN BINDING PROTEIN B.		
GN	TBPB.		
OS	Moraxella catarrhalis.		
OC	Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae; Moraxella		
OX	NCBI.TaxId:480;		
RN	SEQUENCE FROM N.A.		
RC	STRAIN-IES-1.		
RX	MEDLINE-98380363; PubMed-9712766;		
RA	Myers L.E., Yang Y.P., Du R.P., Wang Q., Harkness R.E.,		
RT	Schryvers A.B., Klein M.H., Loomore S.M.;		
RT	"the transferrin binding protein B of Moraxella catarrhalis elicits		

RT bactericidal antibodies and is a potential vaccine antigen.
RL Infect. Immun. 66:4183-4192(1998).
DR EMBL; AF039313; AAC34278.1; -.
DR InterPro: IPR001677; Transfein_bind.
DR Pfam: PF01298; Lipoprotein_5; 1.
SQ SEQUENCE 713 AA; 76811 MW; A92445A61E9A05BD CRC64;

Query Match	3.3%	Score 123.5;	DB 2;	Length 713;
Best Local Similarity	19.48%	Pred. No. 9.6;		
Matches 133; Conservative	76;	Mismatches 269;	Indels 167;	Gaps 24;

QY	119	IPILPSVKYRLKIKYIVDKYEGOLKGTIDMELTKRAEGVCEGLAKKNADENQCLIVE	178
Db	156	IPEDKNILEYLLKKSSEVSKFEAGKGGIENNTRLTHLDLSEQKEAVYKRALDNL---	2111
QY	179	PVYSNNEXFLY---YDVFCKDVRWVFAPSSVSGFGGDDTDMWMPRHGTGDFSVFYVGA	2355
Db	212	TQFAQEKKELEIENAHDKKSAR-----	234
QY	236	DNRPAEYSKDKPKYPVYFAAVSMQGYKADYAMTIGF-----PGST	2777
Db	235	-NRDLEYVKSGFNY-----LSGTATADHDKNTVNGYGYALYYKSGEAKELPQTS	284
QY	278	DRYLISMG--VEDRIENENNRIEVRGI--KGIYKEMASADQ-ATRI-----KASK	3255
Db	285	AYKKYQWQFMTDITLDMKYT--DLPGIARTQIRSLVSTDEVATVTLTLDKNNKPSDNGA	3411
QY	336	YASQANMYKNSIG-----MNRGLARLDYIGRRRAEERAFADWIRKNGKSAVYGDVLS	378
Db	342	YOHSEFPDVNPADKKIKGKLISNLSGSAVYAKERYKIENDIHGNRRPGSA-----TA	394
QY	379	LEKAYKEGAKANREKTY-LSETLFGEGTEVVRFAQFANALATNPDAHAGILKSLDDKYDY	437
Db	395	SDKA--EDSKTOHPPTSDATKMLGSGFYGPKEELACKFLTDDKLLGVGAKRDKEYKT	452
QY	438	LPSLDRKVLPLMLDIVRRIRPADLPRDIEKVID--KFKKGDPKYYAD--PYFDGSVYVYSD	495
Db	453	EALIDATVYL-----GTENNTKATTFTPTTKQLDNFGNAKKLYGST	495
QY	496	KFHAMLKSMDEKEFAKAIENKPAVELSKSVYIAARAIOADAMANAVALER--GKRLFFAG	5533
Db	496	VINLUSTDATKNEFLKFKTKD-----KPTSATNKAGBT-LMWDEVIYVYTGKNFEYTK	548
QY	554	LREATPGRA---LPSDANFTMRSSYSIGKYEPODGAWNYHT---TGCGVLEKQDPKS	606
Db	549	FGELSEVGSHSVLEFGERTATTGKERAPVTTGKAKYLLGNWGYITGAGTGKSFNEADP---	605
QY	607	DEFAVGEMILDLFRKNKGRYAENGOLHIAFLSNNDITG-----	6455
Db	606	---IADDIDIFERKSVKGLTGTGRIDPVFNIGETIAGNGWTGKASTYKADAGGYKIDS	661
QY	646	GNSGSPVFDKNGRLGLAFADNWEAMSGDIEFEPD	680
Db	662	SSTGSKIYIENAEVYCGFYGPANAMMGSSFTHDPD	696
RESULT 8			
ID	Q9X087	PRELIMINARY;	PRT; 1289 AA.
AC	Q9X087;		
DT	01-NOV-1999	(Tremblrel, 12, Created)	
DT	01-NOV-1999	(Tremblrel, 12, Last sequence update)	
DT	01-MAY-2000	(Tremblrel, 13, Last annotation update)	
DE	HYDROTHERMAL 148.5 KDA PROTEIN.		
GN	TMO992.		
OS	Thermotoga maritima.		
OC	Bacteria; Thermotogales; Thermotoga.		
OX	NCBI_TaxID=2336;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=MSB / DSM 3109;		
RX	MEDLINE=99287316; PubMed=10360571;		
RA	Nelson K.E., Clayton R.A., Gill S.R., Gwin M.L., Dodson R.J.,		

RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.,
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of *Thermotoga maritima*.";
RL Nature 399:323-329(1999).
DR EMBL: AE001761; AAD36071.1; -
DR TIGR: TM0992; -
KW Hypothetical protein: Complete proteome.
SQ SEQUENCE 1289 AA; 148491 MW; 82E5081AD98ADCAA CRC64;

Query Match 3.3%; Score 123; DB 16; Length 1289;
Best Local Similarity 18.1%; Pred. No. 26;
Matches 125; Conservative 111; Mismatches 256; Indels 200; Gaps 29;

OY 102 DHDYRDGVSTKMGEEELPIPLSVKYLKIVKVDKVGOLKLTID----- 148
DB 22 DSDALR--YRLWTISE-----LNKSYLKAIONIVG-YNGLEKSGSDFLETLYSNDPLL 72
OY 149 -----EMERLKAQEVCOELAKKENADENOICIVEPFYSNNEFL----- 188
DB 73 ARYKESQAIKRYQARREGASNEIILKQLAYQEARLREKKOPIALIESMLYNSIDYL 132
OY 189 -----IVYDFKDVRFVAPSPSSVGKFGGDTDNMMMPRATGDSV-FRYVA 233
DB 133 ASETLKSLSKTANKIILDVLMKMKELKPLE-GDGFVGIVKNAILLQIDISTYKIFIDYC 191
OY 234 ----GADNRPAEYKDNKPYKPYFAVAVSMOGYKADYMTGTFPGSTDRYLT--SMGVE 287
DB 192 KANYATEKRIAY-----WMKTYF---VEPFESSEKKVL-----DEVLTAKSDEIK 235
OY 288 DRIENENNRIEVRIGKOG-----IMKEA--MSADQATR-----IKYASYAQA 330
DB 236 DILKRLRLRLLEIYSKGNALAEKIEETAKKSAESILKNIETPSLIVELFTKYINYV 295
OY 331 NY--WKNSIGMNGRL--ARLDVIGRKRAE-----ERAFADWIRKNGSAVYGVLSL 379
DB 296 DQOIFINDIASNEVYFIKQIRELVGNDENEINRCYFDRAYFLSKKAKKAINDVGEKT 355
OY 380 EKAYEGAKANKEMTYLSTLFGTEVYVFAQFANALATNPRAHAGILKSLDK----- 433
DB 356 QGGSTKGOKSEEDVYIPOEVAKAEIV-----LAKDPKVASAVONITRTIEEND 407
OY 434 --YKDYLPSLDR-----KVLPAWDIVRRRIADK 461
DB 408 LSLRALPPLNSAKRLIETFGOLENNEITLGEYNEIRIYSVTTGOLNTAKKTIANNL 467
OY 462 LPDIFKNVI-----DKKFKGDTKKYADVF-----DKSVVPSDKFHAMLS 503
DB 468 ESEYHKGISQSEYSEBKQLQDKDVEGNIRQLNDPFLYKHOELERQORFKKLSQLVAT 527
OY 504 MKKEFAKAIENDPAVELSKVIAAARAIQADAMANAIAIEGKRLF--PAGLEBMYG 560
DB 528 LDQGGSEMKL-----LQDSDLA-----KFNQFLKLYQKIGQAVWSAFSFSFSEMMG 574
OY 561 RALPDSANETMMSYGSIKYEPDQAVMYHTTG-----KGVLEKODPKSDEF-A 610
DB 575 FINGIGKAPFKELNSKSPSFLQNGGWTLENLAIQEDERLTRLLEKVDKLSSEIKA 634
OY 611 VOENILDFRTKNYG--RYAENGOLHIAFLSN 640
DB 635 LVESFPRCYTDEDGLVYSKNAMLIEEFKSS 666

RESULT 9
O92B47 PRELIMINARY; PRT; 742 AA.
AC O92B47;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)

DE CPA.
GN CPA.
OS Streptococcus pyogenes.
OC Bacteria: Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CS101;
RA Podbielski A., Moischnik M., Leonard B.A.B., Schmidt K.H.;
RT "Characterization of ure, a global negative regulator gene in group A
RT streptococci.";
RL Mol. Microbiol. 31:0-0(1999).
DR EMBL: U49397; AAC97148.1; -
SQ SEQUENCE 742 AA; 83683 MW; 97A1FF44B4ECB944 CRC64;

Query Match 3.3%; Score 122.5; DB 2; Length 742;
Best Local Similarity 21.2%; Pred. No. 12;
Matches 133; Conservative 62; Mismatches 212; Indels 221; Gaps 31;

OY 206 SVKFGGDTDNMMMPRHHTGDSVFRVYAGADNRPAEYKDNKPYKPYFAVSMOG---Y 262
DB 40 STRAFGEQGS--VNRQSSIDYPMY-GYDSYPRGY-PDYSPLKTYHNLKYNLEGSXDY 95
OY 263 KADYAMTIGPFGSTDRILTSKGVEDRIENEN-----NPRIEVNGIKOGIKKEMASADQ 316
DB 96 QAYCFNLTKHPFSKSDSVRSQWYKKLEGTENENFIKLADPRIEDQLOONI----- 146
OY 317 ARIYVASYAASANYWKNSIGMNGRLARLDY-----GRKRAEERAFADWIRK 365
DB 147 -LRILY-----NGYPNRNNGIMKGIIDLPLNALIVTQNAIWTDSAQINPDSFTEANS 198
OY 366 NKSAYVGDVLSLEKAYEGAKANREMYLSETLFGTEVYVFAQFANALATNPDAHAG 425
DB 199 NG---INNOQLGMKMKALKELIDPVLGSKYSNKTPSG-----YRLN 236
OY 426 ILKSLDKRYK-----DYLPSLDRK---VLPAMLD-----IVRRRIAD-----KL-- 462
DB 237 VESHDKPFQNLISAIEYVDTPPKGEEPPAKTEKTSVIRKYAAGDSKLLLEGATLKLSQ 296
OY 463 -----PDIEFNVIDKKFKKGDTKKYADVFDDK 488
DB 297 IEGSGFQEKDFOSNLGETVELPNCGYLTETSSPDGKIAPRIFRKENKVV--FIYOK 354
OY 489 S-----VVPYSKFFHMLKSMDE-----KFAKAIENDPAVELSKSVIA- 527
DB 355 DGSQVENPKKEVAEPYS--VEAYNDFMDEVLSCGTPYCKFYATNKKSSQVYVCFNAD 412
OY 528 -----AARAIQADAMAN--AYAIKGRLEFAGIREMYPGRALPDSAN--FTWR 572
DB 413 LHSPPDSYDSGETINPDISTMKEKRYTHAGSDLEKVALR-----PROTNPEDFLKH 464
OY 573 MSYGSTIKGYEPDQAVMYNHTTGKGVLEKODPKSDEFANQ--ENILDLFRTKNYRYAVEN 630
DB 465 IKKVIKGYK-KKGSYN-----GLTETQFRAAQALAYTTDSADLKTGLTY-----NN 513
OY 631 GOLHIAFLSNND-----ITGNGSGS-----PVFDKNGRLIGLAFGNMGA 670
DB 514 GKGYGFESMEDKTYLAIVKELITTYAONGSAPOLTYMLDFFVPRNNSDQSLIG----- 564
OY 671 MSGDIEFPDLORTISVDIRVYLEMIDK 698
DB 565 ----TECHPD-----DLVDVIRMEDK 581

RESULT 10
O92B35 PRELIMINARY; PRT; 1571 AA.
AC O92B35;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE LIM1716 PROTEIN.

```

GN L1N1716.
OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Listeria.
OX NCBI_TaxID=1642;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CLIP 11262 / SEROVAR 6A;
RX PubMed=11679669;
RA Glaeser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Bloecher H., Brandt P., Chakraborty T.,
RA Charbit A., Chetoui A., Couve E., de Darvar A., Depoux P.,
RA Domann E., Dominguez-Bernal G., Duchand E., Durant L., Dussuguet O.,
RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Haut J., Jackson D.,
RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst J., Kurapkhat G.,
RA Madeno E., Maitournam A., Mata Vicente J., Ng E., Nedjati H.,
RA Nordisler G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schueter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT "Comparative genomes of Listeria species.";
RL Science 294:849-852(2001).
DR EMBL: AL596169; CAC96947.1; -.
DR Listlist: LIND1716; -.
KW Complete proteome.
SQ SEQUENCE 1571 AA; 167910 MW; B3B69809D9581F6 CRC64;

Query Match 3.3%; Score 122.5; DB 16; Length 1571;
Best Local Similarity 18.7%; Pred. No. 39;
Matches 138; Conservative 104; Mismatches 308; Indels 187; Gaps 31;

OY 11 GAALLGASGVAKAD-----KGMWLLNELNGENLDMRMLGFTLPDLSLYSDKPS 61
DB 559 GGAVLPGGKGGKGVKLFSTATGKMGKFGKL---AGGVGLVGLSATLIGMKNN 614
OY 62 IANAVYFGGCGTGTIVSDGLIFTNHCYGA-----IOSQT 100
DB 615 AGEKGTGGFGSG--MAGGAITGMIAFGVTAIGATGAFGTAIGKELKYQKEG 672
OY 101 VHDYLRDGEVST--MGEELPIGLSVKLYKTVK-----YT 136
DB 673 KILDKFKTKMGKLSIAEHPILGANININKTIDAAGKIKAGDTHKTVNASKSIYA 732
OY 137 D--KVGOLKGTIDEMERLKAQEVCOELAKKENADENL-----CIYEPPYSNEF 187
DB 733 DPLKIDASSGYSKD-----SAKANNVEYLNEQKMODSRRELINVSRRITKELENN--- 784
OY 188 LIYVDVFKDVRVAFAPSSVGFSGDPTDMMWMPRHGTGDFSVFVYAGA---DNRAEYS 243
DB 785 IKTYDKMSD-QLIAATEKKSANKND-----WDKLVALGAVSKDTPASAKKHTNNEASTIN 838
OY 244 KDKPKPKPYFAAVSMQGIKADDIAMTIGPGSGTDRIYLSWGVEDRIENENNRIEVRGI 303
DB 839 KADIKANNELKLEREFYK-----GOEETIRNAEN---RINAL 874
OY 304 KOGIWE--AMSADOATRIKAYKASQASANYMKNSTIGMNGRLARL--VYGRKAEERAF 359
DB 875 KQASSEKRLITDEKELO-----RMETTTLEFRSEBRST 911
OY 360 ADMIRKNGSAVYDVLSSLEKAYKGAANREMTYLSFTLFGCTEVRPAQFANLAT 418
DB 912 EKEVRK-----IEEKQKKEAIALATASAKQKILGNLNSKKMSAKKAASV 959
OY 419 ---NPOAHGILSLDDKDKDYLPISDRKVLPMMLDIVRRIIPADKLPIFFKANIDKFK 475
DB 960 VKNSAARDAASVANKREYKOTKRIIDK-----RFVYGEISEEYODALKNA--KKKK 1011
OY 476 GDRKKTADFEVDSVYPYSDKFAHMLKSMQKEFAKAIKEDPAVELSKSVIAAARAIQAD 535
DB 1012 NGVYKAEKHHNDVREAOQAAKGLREVDME--GETLTK--WEQKKGTVETFPNSTKDA 1068
OY 536 AMANAYAIKGRKLPFAGLEMTYGRALPSDANFTMRMSYGSIKGYEPQDQAMVNYHT-- 593

```

```

DB 1069 ALGKNNELSSATVNIYFSGMKDT-ALRVFDSFKTNLKYVINNVITGVN-KVLSFENIKTIP 1126
OY 594 -TGKYLEKODPKSDEFVAQENILDFRTKNYGRVANKOLHIAFNSNDITG----GNS 648
DB 1127 VLGNGSIGSS-----QENSLSAKDKRYTHSTQSQNLAMNYGNSNNAQIMAGEE 1177
OY 649 GSPV-FDKN---GRLIG 661
DB 1178 GFEIAYKNNSAQARILIG 1194

RESULT 11
ID 099FL1 PRELIMINARY; PRT; 2204 AA.
AC 099FL1;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE POLYPROTEIN
OS Porcine teschovirus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Teschovirus.
OX NCBI_TaxID=118140;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VIR 460/88;
RX MEDLINE=21105197; PubMed=11160660;
RA Zell R., Dauber M., Krumholz A., Henke A., Birch-Hirschfeld E.,
RA Stelzner A., Prager D., Wurm R.;
RT "Porcine teschoviruses comprise at least eleven distinct serotypes:
RT Molecular and Evolutionary Aspects.";
RL J. Virol. 75:1620-1631(2001).
DR EMBL: AF296095; AAK12389.1; -.
DR HSSP; Q88590; 1TMF.
DR MEROPS; C03.UPC; -.
DR InterPro: IPR004004; Calici_pol_hel.
DR InterPro: IPR001676; Rhv.
DR InterPro: IPR000605; RNA_helicase.
DR InterPro: IPR001205; RNA_pol_P3D.
DR Pfam; PF00073; rhv; 2.
DR Pfam; PF00680; RNA_dep_RNA_pol; 1.
DR Pfam; PF00910; RNA_helicase; 1.
DR PRINTS; PR00918; CALICVIRUSNS.
SO SEQUENCE 2204 AA; 247225 MW; FD6B5B5FC6B93BD8 CRC64;

Query Match 3.3%; Score 122.5; DB 12; Length 2204;
Best Local Similarity 19.5%; Pred. No. 66;
Matches 137; Conservative 100; Mismatches 263; Indels 203; Gaps 30;

OY 19 SGYAKADKGMWLLNELNGENLDMRMLGFTLPDLSLYSPDKPSIANAVYIFGGCGTGTIV 78
DB 572 SEMSEAMOGHYLYYDYGVS-----SVDLVIYVSPYDPSPO-----TPYTK 614
OY 79 SDGLIFTNHCYGAIOSQSTVDHLYLRDGEVSRMGELPIGLSVKLYKRYKIVYTK 138
DB 615 TDPPYISTN-----HPYIVDPYMSGYTTIQNTNLASPPSPSQADLVEYFVG 665
OY 139 VEGOLKGT-----DEMERLKAQEVCOELAKKENADENOLCI--VEPFSN----- 183
DB 666 EDVVLKGPSNIFASFOGDTNPSLPQAPETGCDRAMTTHASNENPIPLPVEISGSSRAVFW 725
OY 184 NEYF-----LIVYDVFQD-----VRWVFAPSSV 207
DB 726 DRYFHISTTGTGNTIKPQYRLNITSLRRIILROSLHATYLRGLSIRVAFASPANV 785
OY 208 GKFGDPTDMMWMPRHGTGDFSVFVYAGADNRPAEYSKN-----KPYK-PVYFAAVS 258
DB 786 ECHDGLTVNLVLYPPSGMNININNSTSIEQADSDVNDALPSFTWKAQTFVFTCSVP 845
OY 259 MCGYKADDIAMTIGFPGSGTDRIYLSWGVEDRIENENNRIEVRGIGKOGIWEKMSADOAT 318
DB 846 YVSF-----TSVLVSVSGFETSYQVPR----- 869

```

```

QY 319 RIKYASKYASQANSYKNSIGNRGLALDVIYGRKRAEERAPADWIRKN-----GKSAYV 372
D 870 -----SNOIFODFGGML-VLRSSSTTELTFESLWREFKIRLFCPRGIH 913
QY 373 GDVLSLEKAYKEGAKNARE-----MTYLSLETFEGCTEVNFAQANMALATNPMAHAIL 427
D 914 DS--AAFSKGHPEDSSGSNERLSATMTLS--FQGPATNFS-----LL 953
QY 428 KSLDDKYDYPLSLDRKYLPAMLDIVRRIRPA-DKLPDIFKNVIDKFKGDTKKYADVFV 486
D 954 KQAGVEENPGPSLS-KLYQLKD-----PAVDALCEAYDEL-KKFKEGATNLDSVS 1004
QY 487 DKSVPYPSDKF-----HAML--KSMDEKFKAKIEKDPAVELSKSVIA-AAARAQADMA 538
D 1005 GDSENPMLNFKFYLYGAYILAMKSLHDMPTAAVCFIIGSDITAFVYSKLAKHLKFAKT 1064
QY 539 NAYAIK-----GKRLFFGLREMYGRALP--SDANFTMRMSGYSLKGYEPD----- 585
D 1065 DPPPKPRACKKECCCGGNHNPDELNPSESGFWSRFRKGHPQG--PMODVSSLNIL 1123
QY 586 -GAMYNHTTGKGVLEKODPKSDEFAVOENILDFRTKNRGY 627
D 1124 KGAEMVYHOFERVKIKMTWRTAEYVSSDFLQA-QMDYPHY 1165

```

RESULT 12

```

Q92BR7 PRELIMINARY: PRT: 721 AA.
ID 092BR7
AC 092BR7
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE LIN1477.
GN LIN1477.
OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Listeria.
OX NCBI_TaxID=1642;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CLIP 11262 / SEROVAR 6A;
RX PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
BAguero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Cherouani F., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurgat O.,
RA Entian K.-D., Fishi H., Garcia-del Portillo F., Garrido P.,
RA Gattlier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kreft J., Kunz M., Kunst F., Kutapkat G.,
RA Madueno E., Maitouram A., Mata Vicente J., Ng E., Nedjari H.,
RA Nodules G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehlund J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001);
DR EMBL: AL596168; CAC96708.1; -
DR ListList; LIN01477; -
KM Complete proteome.
SQ SEQUENCE 721 AA; 79910 MW; D0A0D8167AA6837E CRC64;

```

```

Query Match 3.3%; Score 122; DB 16; Length 721;
Best Local Similarity 22.1%; Pred. No. 12;
Matches 112; Conservative 69; Mismatches 195; Indels 130; Gaps 24;

```

```

QY 277 TDRYLTG-WGYEDRTENNNPRIEVRIGKQ-----IMKEANSADQATRI----- 320
D 123 TDRDKDYWILTHQTESNRLUSAKQALESSKAYIQVEKYNDNDIALTTEDLKVAITY 182
QY 321 -KYASKYASQANSYKNSIGNRGLALDVIYGRKRAEERAPADWIRKNKSAVYGVLSL 379
D 183 KMTLTGYAMFESVYKKNKVTDEELARVSENNDLPGVTTTDMNKKYYTYDETLRLISGV 242
QY 360 EKAYEGAKANREMYTLETLFGTEVVR--AOANALATNP-----DAHAGILKS 429

```

```

D 243 STA-KEGLPKDAEYLSQGYSRNDRGKSYLEAOYESVLGSKSQSESVLDSKGNIIET 301
QY 430 LD-----DKYKDYPLSLDRKYLPAMLDIVRRIRPADKLPIDIKVNLDKFKGDTKKYA-DF 484
D 302 VSKTEGSKGKDLVLSVDEYFKAVEDILRNKIK-----QGQIYAGSD 343
QY 485 VEDRSV-----PYSDKFAHL--KSMDEKFKAKIEKDPAVELSKSVIAAARAQAD-- 535
D 344 LFDRAFYVAMPYSGEVLALAGQKLNKGEF-----DDSLGTFTTAAVAMGSAVKGSTIL 398
QY 536 -----AMAN-----ATAIEKGRLEF-----AGIREMYGRALPBDAN-----FTYRM 573
D 399 GGINDGAISSNVTFTDPIALKGTRPKSSWENRKGAGNRPDPVGALEISSNSYMYGVAM 458
QY 574 SYGSIKGYEPQ-----DGAMYNHTTGKGVLE-----EKODPKSDEFAVOENIL 616
D 459 KMGAK-YVNGPLRAPLSTDDMRITYNQGFLGKYGIDLPGEQVYKGGDQITGK-IL 516
QY 617 DLFRKNYGRYAENGQLIHIAFLSNNDITGNSGSPVDPK-----NGRLIGLAFDGNME 669
D 517 DF-----ATGQDYSTPLQMAQYVSTIANGSRIRAPSMVKELRNSTNDVGTATAN-- 570
QY 670 AMSGDIEFEPDLQRTISV--DIRYV 692
D 571 -----EPKVLNKIGVSDNDIKTV 588

```

RESULT 13

```

Q9PPY9 PRELIMINARY: PRT: 743 AA.
ID 09PPY9
AC 09PPY9
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE DNA HELICASE II.
GN UVRD OR UVS01.
OS Ureaplasma parvum (Ureaplasma urealyticum biotype 1).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Ureaplasma.
OX NCBI_TaxID=134821;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SEROVAR 3;
RX MEDLINE=20500219; PubMed=11048724;
RA Glass J.I., Lefkowitz E.J., Glass J.S., Heiner C.R., Chen E.Y.,
RA Cassell G.H.;
RT "The complete sequence of the mucosal pathogen Ureaplasma
urealyticum.";
RL Nature 407:757-762(2000).
DR EMBL: AE002148; AAF30913.1; -
DR HSSP; P56255; 1PJR.
DR InterPro; IPR000212; UvrD-helicase.
DR Pfam; PF00580; UvrD-helicase; 1.
KM Complete proteome.
SQ SEQUENCE 743 AA; 86740 MW; 9D412574673D71E7 CRC64;

```

```

Query Match 3.3%; Score 122; DB 16; Length 743;
Best Local Similarity 17.7%; Pred. No. 13;
Matches 143; Conservative 129; Mismatches 201; Indels 256; Gaps 40;

```

```

QY 10 LGAALLGASGVAKAD-----KGMWLNELNQNEDRMRELQFTL--PLDSLYSFDRPSI 62
D 22 LGPOLVAGACTGKTSVTLRIAYLTRE--KNHPSRLGFTFTNKADEM-----KERV 74
QY 63 ANAYVI-----FGGCGTGTIVSDGCLIFTNHHCYGGAISQSTVVDHDIRDQFVSRSTM 115
D 75 GKTIGVSIPYLSYFHSMC--VKILQODIHYLYNH--NNIKITDIDQEVY-----LKEI 124
QY 116 GEELPIPLSVYKRYKIVYDKVEGOLGKTD-----EMERLRKAQEV----- 159
D 125 FQDLNTEKKS-QVITKIITISKYKKNKFFDQNDMLNEKNHXYLELVLDLNDQARLVDIYKI 183

```

```

QY 160 -COELAKKENADENOLC-----IVEPYSNNETFLIVDFKVR----- 198
DB 184 YCQRCRKLNVDFDILLNLHLFIEPEYLEKWKQKFDVILV--DEFQOTNKKIQYDLIS 241
QY 199 -----WYAPSSSGKFGGDTDNMMKPHRTHGDFSVFRVYAGADNRPAYSKDN-K 247
DB 242 LLATKHQNLFFVGDPPQMTISFRG-AEOWIINNFSQNFK-----NKKVILKNTNYR 291
QY 248 PYKPYEFAVSMQGYKADYAMTIGFPGSDRYLTSNGVDRITENENNRIEVRG--IK 304
DB 292 STQPLILNTANRLDANNNN-----KNNLTAVNT-----NDNNLPLYLKGONPID 336
QY 305 QGIV-----KEAMSDQATRIKYSKYSQASANYWKNISGMNGLARLDVIGRRAR 355
DB 337 EAMWIRAKIRELEEGPRANQIVLFRSNHVSRTIO-----GMMSISITYLLGSKFY 391
QY 356 ERA-FADWIRKKNKSAVGYDVLSSLEKAYEGAKANREMYL-----SEPLFGCTEYVR 409
DB 392 ERKEIKDMT-----AYLKVVNDLD-----ELSEFLRIINTPRRRAIGPTTEHY 433
QY 410 ACPA-----NALATNPDAH-----AGILKSLDDKYYKDYLPSLDRKYLPA 448
DB 434 KHAIAINNNEFLERLAEVENHNNIINTOKKNIINPVNLKEIDEMEDL-----KIHE 486
QY 449 MLDIVRRRI-----PADKLPDIFKNVLDKKEGDTKKYADVPFKSVVPSYSDKPHAML 501
DB 487 ILELIYKVVYEAYLLENKAEKIDNVEE--LKRAKMYVDRHDPDTINDYINSTALYL 544
QY 502 KSMDEK-----FAAITEKDP--AVELSKSVYIAARAIQADAMANAAYIEKGRKL 549
DB 545 NKDGKSKENVLLMTVHNSKGLYEENVEVAGWMEGLIPSDRAINDPIK--GVEEERRI 601
QY 550 FFAGL-----REMYGRALPSPDANTMMSYSGISYEPDQAWY 589
DB 602 AYALIRAKKNLYISACCDPLARKQVFSKRIINEGFNNLKIINSFKNKPED----- 657
QY 590 NYHTTGKGVLEKODPKS--DEFAVQENIIDLFRTKNGYRAENGOLHIAELSNNDITGYN 647
DB 658 ---MPLKSFILKQOEERSWFSKQKQKEVEDNIFYO-----TKMND----- 693
QY 648 SGSPVFDKNGRLGLAF--DGNMEAMSGDI 675
DB 694 ----FEIGERIVHTSFGDGVITIGDGI 717

RESULT 14
O97S45 PRELIMINARY; PRT; 883 AA.
AC 097S45;
DT 01-OCT-2001 (Tremblrel. 18, Created)
DT 01-OCT-2001 (Tremblrel. 18, last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, last annotation update)
DE VALYI-TRNA SYNTHETASE.
GN SP0568.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-TIGR4;
RX MEDLINE=21357209; PubMed=11463916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., Deboy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwin M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzaple E., Khouli H., Wolf A.M., Uterback T.R., Hansen C.L.,
RA McDonald L.A., Feilblyum T.V., Anguoli S., Dickinson T., Hickey E.R.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RL pneumoniae.";
Science 293:498-506(2001).

```

```

DR EMBL: AE007366; AK74724.1; -.
DR TIGR: SP0568; -.
DR InterPro: IPR002300; tRNA-synL_1a.
DR InterPro: IPR001412; tRNA-synL_1.
DR InterPro: IPR002301; tRNA-synL_1le.
DR InterPro: IPR002303; tRNA-synL_val.
DR Pfam: PF00133; tRNA-synL_1; 1.
DR PRINTS: PR00986; TRNASYNTHVAL.
DR PROSITE: PS00178; AA-TRNA_LIGASE_L1.
DR AMINOacyl-TRNA synthetase; Complete proteome.
KW SEQUENCE 883 AA; 100775 MW; F5009144F6A82597 CRC64;

Query Match 3.3%; Score 122; DB 16; Length 883;
Best local Similarity 20.0%; Pred. No. 17;
Matches 102; Conservative 66; Mismatches 167; Indels 176; Gaps 24;

QY 192 DVF-----KQVNVFAPSSVGKFGGDTDNMMKPHRTHGDFSVFRVYAGADNRPAYS 243
DB 26 DVFKPSGDQKAKPYSIYIIPPNTGKL-----HLG-----HA 57
QY 244 KDKPKPYFAVFAVSMQGYKADYAMTIGFPGSDRYL-TSMGVEDRIENENNRIEVRG 302
DB 58 WDT-TLQDIIIRKRMQGF-----TLMLPGMDHAGIATQAKVEERLREGITRRYDLGR 110
QY 303 ----IKQIWKAMSDQATRIKYA--SKYASANYWKNISGMNGLARLDVIGRRKRAEER 357
DB 111 ESFLTWKWKD-----EYATIKQNGKMGSLVDYSNERTLDEGIS-----KAVRK 158
QY 358 APADWIRK-----NGKSAVYGD-----VLSSLEKAYK-----EGAKANREMT 394
DB 159 VFVNLKKGWIVYGEFLINNDPAARTALSDIEVIHDKVEGAFYHMYMLEDGSRALAEVAR 218
QY 395 YLSEETLFGCEVVRPQAFALATNPDAHNLKSLDDKDKDYLPS-----LDRKYLPA 449
DB 219 TREIWMGDV-----AVAVNPE-----DPRKIDLGKVVILPIANKKLPIY 259
QY 450 LD-----IVRRIRPADKLPDIFKNVLDKKEGDTKKYADVPFKSVVPSYSDK 496
DB 260 GDHAPDELGTGVKITPAIDPNDPLVGGQRNRLQVNVMMDDGTMLAEFGSM--DR 316
QY 497 FHMLKSMDEKFAKAIKDPRAVELSKSVYIAARA-----IQAD 355
DB 317 FEAR-----KAVVAKLEELGALVKIEKRVASHVSERTGVVVEPRLSTQWFMVMDLAKN 371
QY 536 AMANAYAIKGRFLFRAGLREMYPGRALPS-----DANTMMSYSGISYEPDQ 586
DB 372 AIANQDTEK-----VEYPPRPNDTEFLQWMEVHDWISROLWGH-----QIP 416
QY 587 AMYNYHTTGKGVLEKODPKSDEFAVQENIID 617
DB 417 AMYN--ADGEMVYGEAEPEGDGTQDEDDVID 445

RESULT 15
O99LP5 PRELIMINARY; PRT; 1208 AA.
AC 099LP5;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, last annotation update)
DE PUTATIVE PHOSPHATIDYLINOSITOL 4-KINASE.
GN D9PRL1.7.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=20181732; PubMed=10715324;
RX Tarchini R., Biddle P., Wineland R., Tingey S., Rafalski A.;
RT "The complete sequence of 340 kb of DNA around the rice Adh1-adh2
RL region reveals interrupted colinearity with maize chromosome 4.";

```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 17, 2002, 22:41:11 : Search time 1724 Seconds

(without alignments)
5574.149 Million cell updates/sec

Title: US-10-008-355-2
Perfect score: 3719
Sequence: 1 MQMKLSILGALLGASG.....LFMDKMGCPRLIGELKLI 712

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 13736207 seqs, 6748477542 residues
Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame_plus_p2n.model -DEV=xlh
-O=/cgn2.1/USPTO/spool/US10008355/runat_16102002_121153_27707/app-query.fasta.1.903
-DB=EST -QFMT=fastap -SUFFIX=1st -MINMATCH=0.1 -LOOPEXT=0 -LIST=45
-UNIT=bits -START=1 -END=1 -MATRIX=blomsun62 -TRANS=human40.cdi -LST=45
-DOCCALIGN=200 -THR.SCORE=pcct -THR.MAX=100 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10008355 -CGEN_1_1_1033 -etunat_16102002_121153_27707 -NCPu=6 -ICPu=3
-NO_XLPHY -NO_MAP -LARGEQUERY -NEG.SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7
-YGAPOP=10 -YGAPEXT=0.5 -DELop=6 -DELEXt=7

Database : EST.*
1: em_estba.*
2: em_esthum.*
3: em_estlin.*
4: em_estnu.*
5: em_estov.*
6: em_estpl.*
7: em_estlro.*
8: em_hlc.*
9: gb_est1.*
10: gb_est2.*
11: gb_hcc.*
12: gb_gss.*
13: em_gss_hum.*
14: em_gss_inv.*
15: em_gss_pin.*
16: em_gss_vrt.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	418.5	11.3	591	12	BH392840 AG-ND-162
C 2	397	10.7	500	12	BH386991 AG-ND-148

C 3	397	10.7	537	12	BH388186	BH388186 AG-ND-124
C 4	385	10.4	418	12	BH375201	BH375201 AG-ND-173
C 5	315	8.5	697	12	BH375688	BH375688 AG-ND-120
C 6	314	8.4	426	12	BH400391	BH400391 AG-ND-147
C 7	239.5	6.4	398	9	AW774807	AW774807 EST333958
C 8	187.5	5.0	721	12	BH391233	BH391233 AG-ND-141
C 9	184	4.9	706	12	BH394089	BH394089 AG-ND-150
C 10	177	4.8	555	12	BH400866	BH400866 AG-ND-125
C 11	177	4.8	769	12	BH381877	BH381877 AG-ND-155
C 12	160.5	4.3	1005	12	BH164253	BH164253 ENTTC637F
C 13	138	3.7	262	12	BH391877	BH391877 AG-ND-141
C 14	136	3.7	822	12	BH371846	BH371846 AG-ND-162
C 15	107.5	2.9	2349	11	AK012962	AK012962 Mus muscu
C 16	107.5	2.9	3112	11	AK004749	AK004749 Mus muscu
C 17	103	2.8	868	9	AM057494	AM057494 ca04.f04.x
C 18	102	2.7	663	9	AM057482	AM057482 ca04.f04.x
C 19	102	2.7	864	9	AU205006	AU205006 AU205006
C 20	102	2.7	1095	11	BC015158	BC015158 Homo sapi
C 21	101.5	2.7	1012	9	AL545921	AL545921 AL545921
C 22	101.5	2.7	3340	11	AK004800	AK004800 Mus muscu
C 23	101.5	2.7	3369	11	AK011711	AK011711 Mus muscu
C 24	100.5	2.7	651	10	BE266416	BE266416 601193163
C 25	100	2.7	937	10	BE105351	BE105351 601822064
C 26	98.5	2.6	1131	12	CNS06LE4	AL404066 T3 end of
C 27	98	2.6	667	10	B1391728	B1391728 ppp1n.pk0
C 28	98	2.6	1087	10	BE798042	BE798042 601583645
C 29	97.5	2.6	920	9	AL545962	AL545962 AL545962
C 30	97.5	2.6	957	12	AZ680449	AZ680449 ENTHN69TR
C 31	97.5	2.6	2836	11	AK004712	AK004712 Mus muscu
C 32	97	2.6	946	10	BE685895	BE685895 602143193
C 33	96.5	2.6	2232	11	BC016514	BC016514 Homo sapi
C 34	96	2.6	3098	11	AK019547	AK019547 Mus muscu
C 35	95.5	2.6	509	12	AZ930127	AZ930127 474.dh250
C 36	94.5	2.5	583	12	BH231647	BH231647 1006163C0
C 37	94.5	2.5	758	12	BH476716	BH476716 B06X591TR
C 38	94.5	2.5	772	10	BJ152144	BJ152144 B152144
C 39	94.5	2.5	3300	11	AK019499	AK019499 Mus muscu
C 40	94	2.5	787	9	AU140924	AU140924 AU140924
C 41	94	2.5	846	10	BE66089	BE66089 603295505
C 42	94	2.5	896	10	BE583368	BE583368 7-8D-MT P
C 43	94	2.5	2426	11	BC001905	BC001905 Homo sapi
C 44	93.5	2.5	843	12	AZ687053	AZ687053 ENTHO85TR
C 45	93.5	2.5	1132	10	BM452650	BM452650 AGENCOURT

ALIGNMENTS

RESULT 1
BH392840/C
LOCUS
DEFINITION
AG-ND-162P14.TF ND-7TM Anopheles gambiae genomic clone AG-ND-162P14
DNA sequence.
ACCESSION
BH392840
VERSION
BH392840.1 GI:17338981
KEYWORDS
GSS.
SOURCE
ORGANISM
African malaria mosquito.
Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae
; Anopheles.
1 (bases 1 to 591)
REFERENCE
Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.
Direct Submission of BAC-end sequences from Anopheles gambiae
Unpublished (2001)
Other GSSs: AG-ND-162P14.TR
COMMENT
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: b.loftus@igr.org
This clone is from an A. gambiae BAC library (ND-7TM) provided by

F. H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.

Seq primer: M13 For

Class: BAC ends

FEATURES

Source

Location/Qualifiers
1. 591
/organism="Anopheles gambiae"
/strain="PEST"
/db_xref="taxon:7165"
/clone="AG-ND-148L11"
/clone_lib="ND-TAM"
/note="Vector: pECBAC1; Site_1: HindIII"

BASE COUNT 170 a 122 c 105 g 194 t

ORIGIN

Alignment Scores:

Pred. No.: 1.68e-36 Length: 591
Score: 418.50 Matches: 82
Percent Similarity: 64.58% Conservative: 42
Best Local Similarity: 42.71% Mismatches: 63
Query Match: 11.25% Indels: 6
DB: 12 Gaps: 3

US-10-008-355-2 (1-712) x BH392840 (1-591)

QY 105 TYLEUATGASPGLYPHEVALSERATGTHMETGLYGLUGLEUPROILEPROGLYLEU 124
DB 590 CACCTAACCTATGCTCTGGCCAAAGATATATGTAAGAAATATTAATTAAGCGTTA 531
QY 125 SERVALYSTYLEUARGLYLEVALLYSVALTHRASPLYVALGLUGLYLEU 144
DB 530 TCAGTAAGATCTTTGTAGAAATGATGATGCGACAAACGATACCTTAACCTGAAC 471
QY 145 GLY-----IleThraspGluwecGluArgLeuArgLysAlaGlnGluValCysGlnGlu 162
DB 470 AATGATATGAGCGCTGATCACAGAGAAAGCTATATGATGATGCTGAATC-----AAA 418
QY 163 LEUALALYSLYGLUASNALAASPGLUASGln--LeuCysIleValGluProPheTyr 181
DB 417 GCTATTCAGATCGAATACTGAGAAATGAGAAATACACTGATGATGATGATGATGATG 358
QY 182 SERASNAANGLUtyrPheLeuIleValtyrAspValPheLysAspValArgMetValPhe 201
DB 357 AAAGAAATGATGATTTACTTCTGATTCATGATGATGATGATGATGATGATGATGATG 298
QY 202 ALAPROPSERSEVALGLYSPHEGLYASPTHRASPTTRMETTRPROARG 221
DB 297 ACTCGCGCTCTATCTATGTAATATGATGATGATGATGATGATGATGATGATGATGATG 238
QY 222 HASTHRGLYASPPHESEVALPHEARGVALTYRALAGLYALASPSANRPROLAGLU 241
DB 237 CACAGAGAGACTTCTGCTGTTTCGCTGTTATGCTGATTAAGAAAGCAATCCTCTGAA 178
QY 242 TYRSERLYASPSANLYSPROTLYSPROVALTYRPHLEALALVALSERMETGlnGly 261
DB 177 TATTCAGTATACACGCTCTTAAGCCTTAAGCATTCATTAACCAATTTCAATAAAGA 118
QY 262 TYRLYSALASAPPTVALAMETHRILLEGILYSPHEPROGLYSERTHRASPTTRTYRLEU 281
DB 117 AATATGCTGCTGATTTGCAATGATTTGATGATTAATCCTGATGATGATGATGATGATG 58
QY 282 THRSERTPGLYVALGLUASPARGLIEGLUNSGLU 293
DB 57 ACTTCTTCGGAATGCAACAATGATGATGATGATGATGATGATGATGATGATGATGATG 22

RESULT 2
BH386991/c

LOCUS BH386991 500 bp DNA linear GSS 10-DEC-2001
DEFINITION AG-ND-148L11.TR ND-TAM Anopheles gambiae genomic clone AG-ND-148L11
ACCESSION BH386991
VERSION BH386991.1 GI:17333133
KEYWORDS GSS.
SOURCE African malaria mosquito.
ORGANISM Anopheles gambiae

REFERENCE 1 (bases 1 to 500)
AUTHORS Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.
TITLE Direct Submission of BAC-end sequences from Anopheles gambiae
JOURNAL Unpublished (2001)
COMMENT Other GSSs: AG-ND-148L11.TF
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: b.loftus@tigr.org

This clone is from an A. gambiae BAC library (ND-TAM) provided by F. H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.
Seq primer: M13 Rev
Class: BAC ends.

FEATURES

Source

Location/Qualifiers
1. 500
/organism="Anopheles gambiae"
/strain="PEST"
/db_xref="taxon:7165"
/clone="AG-ND-148L11"
/clone_lib="ND-TAM"
/note="Vector: pECBAC1; Site_1: HindIII"

BASE COUNT

143 a 105 c 90 g 162 t

ORIGIN

Alignment Scores:

Pred. No.: 3.37e-34 Length: 500
Score: 397.00 Matches: 71
Percent Similarity: 65.00% Conservative: 33
Best Local Similarity: 44.38% Mismatches: 52
Query Match: 10.67% Indels: 4
DB: 12 Gaps: 2

US-10-008-355-2 (1-712) x BH386991 (1-500)

QY 134 LYSVALTHRASPLYVALGLUGLYLEUVALYSLYGLYIleThraspGluwecGluArgLeu 153
DB 484 CGTATCATCTTCTAAGCAAGATGATGATGATGATGATGATGATGATGATGATGATG 431
QY 154 ARGLYSALGlnGluValCysGlnGluLeuAlaLysGlnGluASNALAASPGLUASGln 173
DB 430 ATCGATGCTGAACCAAGCATATCAGCTGAAACCTGTGAATGCA-----AAATAC 377
QY 174 LEUCYSILEVALIGLUProPheTyrSerASNAANGLUtyrPheLeuIleValtyrAspVal 193
DB 376 ACTGATGAGGAGGATTTCTTAAGAGAAATAGATTACTTCTGATTAATTCACAGAT 317
QY 194 PHELYASPSVALARGMETVALPHEALAPROPSERSEVALGLYLYSPHEGLYGLYASP 213
DB 316 TTCAAAGATGTTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 257
QY 214 THRASPSNTRPKETTRPROARGHISTHRGLYASPPHESEVALPHEARGVALTYRALA 233

Db	256	ACTGATTACTGGGAATGGCCAAAGCACACAGAGACTTCCTCGTTTCCGATTATGC	197
OY	234	GlyAlaAspAsnArgProIleGluThrSerLysAspAsnLysProTyrLysProValYr	253
Db	196	GATTAATAAGTGCAATCCTCGTAGAATATCAGCTAACACGTTCCTTAAAGCTTAAGCAT	137
OY	254	PheAlaIaValSerMetGlnGlyTyrLysAlaAspAsPTyralaMetThrIleGlyPhe	273
Db	136	CATTACCAAATTTCACTCAATMAAGAATAAGACCTGGTGATTTTGCATGTTCTAGATAT	77
OY	274	ProGlySerThrAspArgTyrLeuThrSerTrpGlyValGluAspArgIIleGluAsnGlu	293
Db	76	CTGTGTACAAACAACAGATACCTTACTTCTTTCGAAATGCAACAATAAGTATGACAAAGAT	17
RESULT_3	BH388186/c		
LOCUS	AG-ND-124P3	537 bp	DNA linear GSS 11-DEC-2001
DEFINITION	AG-ND-124P3, TR ND-TAM Anopheles gambiae genomic clone AG-ND-124P3,		
	DNA sequence.		
ACCESSION	BH388186		
VERSION	BH388186.1	GI:17334327	
KEYWORDS	GSS.		
SOURCE	African malaria mosquito.		
ORGANISM	Anopheles gambiae		
	Eukaryote; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;		
	Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae		
	; Anophelinae.		
REFERENCE	1 (bases 1 to 537)		
AUTHORS	Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.		
TITLE	Direct Submission of BAC-end sequences from Anopheles gambiae		
JOURNAL	Unpublished (2001)		
COMMENT	Other_GSSs: AG-ND-124P3.TR		

Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjloftus@tigr.org
This clone is from an A. gambiae BAC library (ND-TAM) provided by
F.H. Collins and sequenced by The Institute for Genomic Research
(TIGR). The BAC library was generated from A. gambiae PEST strain
DNA. All DNA was extracted from newly hatched first instar larvae
to minimize the inclusion of DNA from microorganisms that inhabit
the gut. The DNA is derived from mixed sexes of larvae. The BAC
library was constructed at Texas A&M University BAC Center
University, College Station, Texas 77843-2123, USA using a HindIII
partial digest.
Seq primer: M13 For
Class: BAC ends.

FEATURES	source	Location/Qualifiers
	1..537	
		/organism="Anopheles gambiae"
		/strain="PEST"
		/db_xref="taxon:7165"
		/clone="AG-ND-124P3"
		/clone_1b="ND-TAM"
		/note="Vector: pECBAC1; Site_1: HindIII"
BASE COUNT	153 a	112 c 94 g 178 t
ORIGIN		

Alignment Scores:	
Pred. No.:	3,76e-34
Score:	397.00
Percent Similarity:	65.00%
Best Local Similarity:	44.38%
Query Match:	10.67%
DB:	12
Length:	537
Matches:	71
Conservative:	33
Mismatches:	52
Indels:	
Gaps:	2

US-10-008-355-2 (1-712) x BH388186 (1-537)

QY 134 LysValThrAspLysValGluGlyGlnLeuLysGlyIleThrAspGluMetGluArgLeu 153
:::|||||::: ::: |||::: ::: :::

D	489	CGTATCACTTCCTAAACGTAACAATAGATTATGAGC-----GGTAGTACAGAAAAGACTTT	436
OY	154	ArgylsalaglnGlualValCysGlnGluleuAlalysLysGluasnAlaAspEubnsGln	173
D	435	ATCCATGTGGTGAATAACAAAGCATATCCAGCTTGAAAACTCTGGAGAATGGA-----AAATAC	382
OY	174	LeucylstleValIglutProphetSerAspnasnglutyRPhelenulleValTytrAspVal	193
D	381	ACTGTAGGGGGAAGGATTTCTTCAAAGGAATAGTTTTACTCTTGTAITTCAGGAT	322
OY	194	PheUlyaspValArlgmetValPheAlaProProserSerValGlyLysPheGlyLysp	213
D	321	TTCAAAGATGTTCCTGTGAGTACTGCCGCATCATATATGGTAATAATATGAGTGAT	262
OY	214	ThraspsnstrPmetTRproArghisThrGlyaspPheSerValPheArvalTytraLa	233
D	261	ACTATATACCTGGATGCCAACAGCACACAGAGACTTCCTGTGTTTCCGCTTATGCT	202
OY	234	GlylaAspaSnarvProAlaglutySerLysAspaSnlsPvrotyLysPvroValtyr	253
D	201	GATAAAAATGGCAATCCTGCTGAAATATTCACCTAACACGTTCTTTAAACCTTAAGCAT	142
OY	254	pheaIaAlaValserMetGlnGlytYLyslaAspasPrtyraLaMetThrIleoglyPhe	273
D	141	CATTACCAATTCACACTAAAGAAATAGCCGTGATTTTGCATGATTGTAGGARAT	82
OY	274	ProGlySerThrAspArgTyrTyleuthSerterPoglValaGluaSparYlgIeLuasnGlu	293
D	81	CCTGTGTCCAAACAAACAGATTACCTTCTTTCGGAATCGAAACAATGTAGGCAAAAGAT	22
RESULT 4	BH375201	418 bp DNA Linear GSS 10-DEC-2001	
LOCUS	BH375201		
DEFINITION	Ag-NP-173C3.TR ND-TAM Anopheles gambiae genomic clone Ag-NP-173C3,		
ACCESSION	BH375201		
VERSION	BH375201.1 GI:1731343		
KEYWORDS	GSS.		
SOURCE	African malaria mosquito.		
ORGANISM	Anopheles gambiae		
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;		
	Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae		
	; Anophelinae.		
REFERENCE	1 (bases 1 to 418)		
AUTHORS	Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.		
TITLE	Direct Submission of BAC-end sequences from Anopheles gambiae		
JOURNAL	Unpublished (2001)		
COMMENT	Contact: Brendan J Loftus		

Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjo1ofus@tigr.org
This clone is from an A. gambiae BAC library (ND-TAM) provided by F. H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.
Seq primer: MJ3 Rev
Class: BAC ends.

```
FEATURES
source
1. .018
Location/Qualifiers
/organism="Anopheles gambiae"
/strain="PEST"
/db_xref="taxon:7165"
/clone="AG-ND-173C3"
/clone_1lb="ND-TAM"
/note="vector: pECBAC1; Site_1: HindIII"
```


DB	148	AAAGATGGCATTTGCACATTCATCAATGGGGGTTTACTGCGCAAAATCGTTTCAAAACAGCGGT	207
OY	83	Leu1lePheThrAsnHisHisCysGlyTyrglyAlaIleGlnSerGlnSerThrValasp	102
DB	208	TTGGTCTTCAACAAACACACACCTGCTGTGTATATACGCGATTCGAAACTTCACATTCACAGAA	267
OY	103	HisAspTyrlLeuArgAspGlyPheValSerArgThrImeGlyGlnGluIleuProIlePro	122
DB	268	CAAAATCATTTTAAAAAATGGTTTGGCGCAAAAAATTCGTCAGAAAGAACTAAACCAAAA	327
OY	123	GlyLeuSerValLysTyrlLeuArgLysIleValLysValThrAspLysValGlnGlyGln	142
DB	328	TCTTTA-----TAGCTTCGTTTTCGTACGATAGATGATGTTCTTAAAGAATT	378
OY	143	LeuLysGlyIleThrAsp	148
DB	379	TTGTCAAAAAGTAAATGAT	396
RESULT 8			
LOCUS	BH391233	721 bp	DNA
DEFINITION	AG-ND-141L2.TF ND-TAM Anopheles gambiae genomic clone AG-ND-141L2,		linear GSS 11-DEC-2001
ACCESSION	BH391233		
VERSION	BH391233.1		
KEYWORDS	GSS.		
SOURCE	African malaria mosquito.		
ORGANISM	Anopheles gambiae		
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;		
	Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae		
	; Anopheles.		
REFERENCE	1 (bases 1 to 721)		
AUTHORS	Shetty,J., Malek,J.J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.		
TITLE	Direct Submission of BAC-end sequences from Anopheles gambiae		
JOURNAL	Unpublished (2001)		
COMMENT	Other_GSSs: AG-ND-141L2.TR		
	Contact: Brendan J Loftus		
	Department of Eukaryotic Genomics		
	The Institute for Genomic Research		
	9712 Medical Center Dr., Rockville, MD 20850, USA		
	Tel.: 301 838 0208		
	Fax: 301 838 3543		
	Email: bjo@fuserg1.fgr.org		
	This clone is from an A. gambiae BAC library (ND-TAM) provided by		
	F.H. Collins and sequenced by The Institute for Genomic Research		
	(TIGR). The BAC library was generated from A. gambiae PEST strain		
	DNA. All DNA was extracted from newly hatched first instar larvae		
	to minimize the inclusion of DNA from microorganisms that inhabit		
	the gut. The DNA is derived from mixed sexes of larvae. The BAC		
	library was constructed at Texas A&M University BAC Center		
	University, College Station, Texas 77843-2123, USA using a HindIII		
	partial digest.		
	Seq primer: M13 For		
	Class: BAC ends.		
FEATURES			
source	Location/Qualifiers		
	1..721		
	/organism="Anopheles gambiae"		
	/strain="PEST"		
	/db_xref="taxon:7165"		
	/clone="AG-ND-141L2"		
	/clone_1id="ND-TAM"		
	/note="Vector: pECBAC1, Site_1: HindIII"		
BASE COUNT	269 a 122 c 147 g 183 t		
ORIGIN			
Alignment Scores:			
Pred. No.:	2.07e-10	Length:	721
Score:	187.50	Matches:	62
Percent Similarity:	43.10%	Conservative:	38
Best local Similarity:	26.72%	Mismatches:	108
Query Match:	5.04%	Indels:	24
DB:	12	Gaps:	5

US-10-008-355-2 (1-712) x BH391233 (1-721)
OY 307 ILETpLySGlUAlMeSerAlaSpGlnAlaThrArgIleTySAlaSerTyS 326
Db 23 GTTATGAAGAACCATATGATTAAGACGATCTACAGATATGCTTACCGCGCTAACTAT 82
OY 327 AlAGInSerAlaAsnTyTThpLySnsSerIleGlyMeAsnArgGlyUeunAlaArgLeu 346
Db 83 GGAAGGTGGCTAACTATTGGAAAAACAGACGGGAACAATTAAGACGGGTTACAAAAC 142
OY 347 AspValIleGlyArgLySArgIaGluGlnAlaGlnAlaPheAlaAspTrpIleArgLySAsn 366
Db 143 GATCATATTGGTCACAAAAAAGAACTGTGAGAAAAAATACCAACATGGCGCTATAAAGT 202
OY 367 GlyLSerAlaValTyTcLySValLeuSerSerLeuGluLySAlaTyTrLySGluGly 386
Db 203 GAAACAAAGCTGTTATTAAGAAATGATATAGCAAAATGACTGATCATACAAACAAT 262
OY 387 AlAlYSAlaAsnArgGluMetThrTyTrLySArgGluThrLeuPheGlyGlyThrGluVal 406
Db 263 TCTAACAGAAAATATGAGAAAATATATGATGTCAGACGTTCCAGAGAAATGCTAAATATACATC 322
OY 407 ValArgPheAlaGlnPheAlaAsnAlaLeuAlaThr-----AsnProAspIla 422
Db 323 AGAAATCTCTTCCAAATCGAGATCGCTTACACAGCTACATGACAGAGATCATCTGCT 382
OY 423 HisAlaGlyIleLeuLySserLeuAsp-----AspLySArgLySAspTyTrLeu 438
Db 383 CAGGACAGCATGAACCTTAATTAAGACAGCTGTTAAGACAGCGCTTAAGAACGGTTAAT 442
OY 439 ProSerLeuAspArgLySValLeuPro-AlaMetLeuSPlIeValAlaArgArgIlePr 458
Db 443 ACTCAACTTGAAGCAGAAATGCTACTCAGATGCTTTGTTTATCAATCAAAAAGGTGGC 502
OY 458 GlnAspLySLeuProAspIlePheLyS-----AsnValIle 470
Db 503 ACCCAATGTTGCTTCTGCACACTGTAAATCTGTAAATGCATCAGACACTTGGCAATTTGC 562
OY 470 eAspLySArgPheLySGlyAspThrLySArgIleAlaAspPheValPheAspLySArgVal 490
Db 563 ACAGCTTCAATTCCTTGCAGACGACGCTTCAATTAATTAACCTCCGACCAATCCAAAGCGC 622
OY 490 lValProTySerAsp-----LysPheHisAlaMetLeuLySserMetAspLy 506
Db 623 TGAATAATTAGCGATGATTAACCTTCAACAAGCTTCTGCAGCGCTATATCCGACAGACAA 682
OY 506 s-----GluLySPhenAlaLySAlaIleGlu 514
Db 683 AGTATTGGCTGAATAATATGCTAAACGTAGTAA 716
RESULT 9
LOCUS BH394089
DEFINITION AG-ND-150P20.TF ND-TAM Anopheles gambiae genomic clone AG-ND-150P20
ACCESSION BH394089
VERSION BH394089.1 GI:17340230
KEYWORDS GSS.
SOURCE African malaria mosquito.
ORGANISM Anopheles gambiae
REFERENCE 1 (bases 1 to 706)
AUTHORS Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.
TITLE Direct Submission of BAC-end sequences from Anopheles gambiae
JOURNAL Unpublished (2001)
COMMENT Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel.: 301 838 0208
Fax: 301 838 3543

Email: b1loftus@tigr.org

This clone is from an A. gambiae BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.
Seq primer: M13 For
Class: BAC ends.

FEATURES
source
location/Qualifiers
1..706
/organism="Anopheles gambiae"
/strain="PEST"
/db_xref="taxon:7165"
/clone="AG-ND-150P20"
/clone_1lb="ND-TAM"
/note="Vector: pECBAC1; Site_1: HindIII"

BASE COUNT 267 a 122 c 140 g 177 t
ORIGIN

Alignment Scores:
Pred. No.: 4.96e-10 Length: 706
Score: 184.00 Matches: 49
Percent Similarity: 47.09% Conservative: 32
Best Local Similarity: 28.49% Mismatches: 82
Query Match: 4.95% Indels: 9
DB: 12 Gaps: 2

US-10-008-355-2 (1-712) x BH394089 (1-706)

OY 307 llettrpysglualametserralaasplnalatrargilelytyralaserlystyr 326

DB 23 GTTATGAAGAAGCATATGATTAAGACGATGCTACAGATAGCTTACGCGCTAATCAT 82

OY 327 lalaginserralaasptlytrpysasnserileglymetasnargglyleualargleu 346

DB 83 GCAAGCTGTGCTAACTATTGGAAAAACAGCCGGACAACTGAAGCGGTTTACAAAAC 142

OY 347 Aspvalilleglyargylarglualagluaralaphelialaspttrpleargylsasn 366

DB 143 GGAACAATTGGTGCACAAAAGAACTGAGAAAAATACCAACATGGGCTGATTAAGCT 202

OY 367 glylysserrala-valtyrglyaspyallleuserleuglulysalatyrylsglucl 386

DB 203 GAAACAAAGCCTGTTATGAAATGTATTGCAAAATCTGATGCTACACAAACAT 262

OY 386 yalalysalasnargglumetthtyrleusercluthrleupheglythrgluva 406

DB 263 TTCTAACAGAAATATCGAGAAAAATTATGTCACAGATTCACAGAAATGCTAAATACAT 322

OY 406 lvalargpheaalaglnpheaalsnralaleualathr-----AsnProAspAl 422

DB 323 CAGAATTCCTTCCGATCGGAGATGCTTAACAAAGCTACATGAAGCAGATGCAATGCG 382

OY 422 ahislaaglyleleuLysSerleuAsp-----AspLytyrTylyAspTylye 438

DB 383 TCAGCAGCTATGAACCAATTAATAGAACACACTGTTAAGCAGGCTTATGAAGGCTTAA 442

OY 438 uProserleuasparglylvalleuproaliametleuaspllevalargalargyllePr 458

DB 443 TACTCACTTGAAGCAAAATGCTATCTCAGATGGCTCTTATATATCAATCAAAAGTAGC 502

OY 458 calaasplysleuProAspIlePheLysAsnVal 469

DB 503 AGCAGATGTGCTCTCGCAACTGTAAATCTGTA 536

RESULT 10
BH400866 555 bp DNA linear GSS 11-DEC-2001
LOCUS BH400866/c
DEFINITION AG-ND-125M4.TR ND-TAM Anopheles gambiae genomic clone AG-ND-125M4,

DNA sequence.

ACCESSION BH400866
VERSION BH400866.1 GI:17347082

KEYWORDS GSS.
SOURCE African malaria mosquito.

ORGANISM Anopheles gambiae
Eukaryote; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae

REFERENCE 1 (bases 1 to 555)
Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.

Direct Submission of BAC-end sequences from Anopheles gambiae
Unpublished (2001)

Other GSSs: AG-ND-125M4.TF
Contact: Brendan J Loftus

Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0208
Fax: 301 838 3543

Email: b1loftus@tigr.org

This clone is from an A. gambiae BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.
Seq primer: M13 Rev
Class: BAC ends.

FEATURES
source
location/Qualifiers
1..555
/organism="Anopheles gambiae"
/strain="PEST"
/db_xref="taxon:7165"
/clone="AG-ND-125M4"
/clone_1lb="ND-TAM"
/note="Vector: pECBAC1; Site_1: HindIII"

BASE COUNT 152 a 105 c 94 g 204 t
ORIGIN

Alignment Scores:
Pred. No.: 2.08e-09 Length: 555
Score: 177.00 Matches: 58
Percent Similarity: 46.82% Conservative: 23
Best Local Similarity: 33.53% Mismatches: 73
Query Match: 4.76% Indels: 21
DB: 12 Gaps: 4

US-10-008-355-2 (1-712) x BH400866 (1-555)

OY 310 GlualametserralaasplnalatrargilelytyralaserlystyrAlagInser 329

DB 545 GAAATATGCGCTAGTAATGCAACCGTATTAATATGCTTAATATGCAATGCAATGCG 486

OY 330 llaasnrytrpplysasnserileglymetasnargglyleualargleuaspyallle 349

DB 485 GCTAAGTATGGGCAAAAATGATGGTAAAGTA-GAAGGTTTAAAAATCTATAT-GCAGTA 428

OY 350 Glyargylsargylarglualarglualaphelialaspttrpleargylsasnglylysser 369

DB 427 GCTAAAAAGCAGCTTATGAAACCACTAGCA-----CAGAAAAACAGAGAGTA 377

OY 370 AlavaltyrglyaspyallleuserleuglulysalatyrylsgluclAlalalysala 389

DB 376 AAAGCACTGTGATTAATTTATGCAATTTG-----TATACGAGCAGGCTCTTAT 326

OY 390 Asnargglumetthtyrleu-----SergluThleupheglyly 403

DB 325 GCATTGAACAATACCTTACTATCTGAGGTTACAAAAAATGCAAGACAACTTGGCTTGGC 266

```

OY 404 ThrGIuValValArgPheAlaGlnPheAlaAsnAlaLeuAlaThrAspProAspAlaHis 423
Db 265 AACATTTCATCATATATTTTCAGAGGTTGAATCCGAAAAAGCTACCCCGGAATCTACT 206
OY 424 AlaGlyIleLeuLysSerLeuAspAspLysTyrLysAspTyrLeuProSerLeuAspArg 443
Db 205 AAGAAACATTAAGAACACTTACTCTTTCTACAGAGATTTATGAAGAGAACTGATCT 146
OY 444 LysValIleLeuProAlaMetLeuAspIleValArgArgArgIleProAlaAspLysLeuPro 463
Db 145 AAGGTACAGCAAAATTTATGTGCTTATATGCACAGAAACCTCCGCTGAATTTTACT 86
OY 464 -----AspIlePheLysAsnVal 469
Db 85 TCAGATTGCTCACTTCACTGATGTAAACAAACCTT 47

RESULT 11
LOCUS BH381877 769 bp DNA linear GSS 10-DEC-2001
DEFINITION AG-ND-155J17.TR ND-TAM Anopheles gambiae genomic clone AG-ND-155J17
ACCESSION BH381877
VERSION BH381877.1 GI:17328019
KEYWORDS GSS.
SOURCE African malaria mosquito.
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae
; Anopheles.
1 (bases 1 to 769)
Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.
Direct Submission of BAC-end sequences from Anopheles gambiae
Unpublished (2001)
Other-GSSs: AG-ND-155J17.TF
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: b.loftus@tigr.org
This clone is from an A. gambiae BAC library (ND-TAM) provided by
F.H. Collins and sequenced by The Institute for Genomic Research
(TIGR). The BAC library was generated from A. gambiae PEST strain
DNA. All DNA was extracted from newly hatched first instar larvae
to minimize the inclusion of DNA from microorganisms that inhabit
the gut. The DNA is derived from mixed sexes of larvae. The BAC
library was constructed at Texas A&M University BAC Center
University, College Station, Texas 77843-2123, USA using a HindIII
partial digest.
Seq primer: M13 Rev
Class: BAC ends.
FEATURES
Location/Qualifiers
source 1..769
/organism="Anopheles gambiae"
/strain="PEST"
/db_xref="taxon:7165"
/clone="AG-ND-155J17"
/clone_lib="ND-TAM"
/note="Vector: pECBAC1, Site_1: HindIII"
BASE COUNT 286 a 124 c 156 g 203 t
ORIGIN
Alignment Scores:
Pred. No.: 3,46e-09 Length: 769
Score: 177.00 Matches: 63
Percent Similarity: 41.84% Conservative: 37
Best Local Similarity: 26.36% Mismatches: 87
Query Match: 4,76% Indels: 53
DB: 12 Gaps: 6
US-10-008-355-2 (1-712) x BH381877 (1-769)

```

```

OY 307 IleTyrPlyGluAlaMetSerAlaAspGlnAlaThrArgIleLysTyrAlaSerLysTyr 326
Db 21 GTTATTGAGAGACATATGATTAAGACGATGCTACAGATTAAGCTTACCGCTCAACTAT 80
OY 327 AlaGlnSerAlaAsnTyrTyrPlyAsnSerIleGlyMetAsnArgGlyLeuAlaArgLeu 346
Db 81 GCAAGTGTGCTCACTATTGTAAGAAAAAGACAGCCGCAACATTTGAAGCGGTTTACAAAAAC 140
OY 347 AspValIleGlyArgLysArgAlaGluGluArgAlaPheAlaAspTyrPileArgLysAsn 366
Db 141 GGAACAATTGCTGACAAAAAAGAGTTGAGAAAAATTTTATTA-TGGCGCTGATAAAGCT 199
OY 367 GlySerAlaValTyrGlyAspValLeuSerLeuGluLysAlaTyrLysGluGly 386
Db 200 GAAACAAAGCTGTTTATGGAATGTAATTAACAAATACTGATGACTACACAACAAT 259
OY 387 AlaLysAlaAsnArgGluMetThr-TyrLeuSerGluThrLeuPheGlyGlyThrGluVal 406
Db 260 TCTAACAGAAATATGCGAAAAATTTATGCTGCACAGTTCCAGAGAAATGCTTAATACATC 319
OY 407 ValArgPheAlaGlnPheAlaAsnAlaLeuAlaThr-----AsnProAspAla 422
Db 320 AGAATTCTTCCAGATCGGATCGCTTACACAGCTACAGACAGATGATGCT 379
OY 423 HisAlaGlyIleLeuLysSerLeuAsp-----AspLysTyrLysAspTyrLeu 438
Db 380 CAGCAGCTGATGAACCTAAATTAGAACAGCTGTAAAGCAGGCTTATGAAGGGTTAAT 439
OY 439 ProSerLeuAspArgLysValLeuProAlaMetLeuAspIleValArgArgIlePro 458
Db 440 ACTCAACTGAAAGCAGAAATGCTATCTCAGATGGCTTATATCAATCAAAATGCA 499
OY 459 AlaAsp-----LysLeuProAspIlePhe 466
Db 500 GCAGATGTGCTGCTTGCACACTGTAATAATCTGTAATGCATCAGAACTTGGCAATTTGCA 559
OY 467 LysAsnValIle----- 470
Db 560 CAGTCTCAATTTTGCAAAGCAGCTTCAGTAATTAACCTCTGAAACAATCCAAAGCGCT 619
OY 471 -----AspLysPheLys-----GlyAspThrLys 479
Db 620 GAAAAATTAGCATGATGATTAACCTTACAGTTTCTCAGGCTATATCGGAGACAAAG 679
OY 480 LysTyr-----AlaAspPheValPheAspLysSerValProTyr 493
Db 680 TATGCTGAAAAATATGCTAAACTGATGAGGTTTCAGAAAGACAGCGTTTATAT 736

RESULT 12
BH164253/c 1005 bp DNA linear GSS 24-SEP-2001
LOCUS ENT163Tf Entamoeba histolytica Sheared DNA Entamoeba histolytica
DEFINITION genomic, DNA sequence.
ACCESSION BH164253 GI:15737691
VERSION BH164253.1 GI:15737691
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica
Eukaryota; Entamoebidae; Entamoeba.
1 (bases 1 to 1005)
Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.
Determination of clone end sequences from Entamoeba histolytica
HMI-IMSS sheared DNA library (2001)
Unpublished (2001)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: b.loftus@tigr.org
Clones are derived from the Entamoeba histolytica HMI-IMSS sheared
DNA library

```


FEATURES

Source

Location/Qualifiers

1.2349

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="MGD:MGI:1895443"

/db_xref="taxon:10090"

/clone="2810401M23"

/clone_jib="RIKEN full-length enriched mouse cDNA library"

/dev_stage="10, 11 days embryo"

/note="data source:SPTR, source key:Q9NYD8, evidence:ISS homolog to CGI-201 PROTEIN putative"

/codon_start=1

/protein_id="BAB28572.1"

/db_xref="GI:12850038"

/translation="MAASTKAKQRIPIKVAKKKAPAEVOITAEQLREARELEL
LPPOOKITDEEELNDYKIKRKKEEDINIKRNVISNVIKQWESLEIKIOPAS
IYERALDVIDYINILMLKYAMEMKROVNAHINMORAITLTPVNOFWYKRYTME
MLGNVAGROYFEREMEMOPEOAMHSYINELRYKVEERARTYERVLVHPAVKM
IKYARFEKHAIFAHARKYERAVEFDEMDDELHYAFKFEENKQEFERVITYK
YALDRISKQAELEFNKNTIFEKFGDRGLIEDITVSRKQYEEEVANPHNDAMF
DYLRLVESDAADAVREYERAIANVPIQERHMKRYLYLVNVALYEADPER
TROYOASLELIPHKETFAKMLYLAQFEIQRKRLPARALGTSICKPKMLFKG
YIELEQLREDRCKLEYKFLERGPENSMIKFAELETIGIERARAYELASIO
PRDMPEVLMKSYIDEIEOEETERTRLVROLQROHVWVTSFAQPELSSGKES
VAKROIIEANKTRNCEKEERMLLESRSFDEFGTVSDKRVDKLMEKVKR
RKVQADDSGDAEMETIDYIFEDANOPNLKLAMATLWKQOQEREAADDPKDI
DESESSP"

BASE COUNT 744 a 451 c 597 g 557 t

ORIGIN

Alignment Scores:

Pred. No.: 1.27 Length: 2349

Score: 107.50 Matches: 115

Percent Similarity: 34.23% Conservative: 89

Best Local Similarity: 19.30% Mismatches: 203

Query Match: 2.89% Indels: 189

DB: 11 Gaps: 32

US-10-008-355-2 (1-712) x AK012962 (1-2349)

QY 62 lleaasnaalavalallephegylgylgys---Thrglylenthvalserasp 80

DB 19 GGGCTGACTGCTGTGGCGGCGGTTGAGAGTCTTTTCACAGTCCGGTGTGGCC 78

QY 81 Gln-glyleuilepethrAsnHisHisCysglYtYrglAlaalleGlnSerGlnSerth 100

DB 79 AGGATCGCTCGAGTTC---CACGGCTCGAGTCATGCGACCTCCAGCGC----- 124

QY 100 rValasphIsaspyrYleuArgAspGlypheValserAgtHnmetGlygluGluLeur 120

DB 125 -----GGCGGGGAAACAGC 138

QY 120 o-ileProGlyleuserValLysTYrLeuArgLysIleValLysValThrAspLysValG 140

DB 139 GCATTCT-----AAAGTGGCCAAAGGTGAAGAAACAAGAGCCG 174

QY 140 lncGlylnleuLysglYlethrAspGluMetGluArgLeuArgLysAlaGlnGluValC 160

DB 175 CGCGCGAAGTGCAG---ATAACGGCCGAG---CAGCTCTCCGAGAGGTAAAGAGAGAG 228

QY 160 ysgGlnGluLeu-----AlaLysLysGluLusnAlaAspGluAsnGlnLeuCysI 176

DB 229 AACTCGAGCTCTCCACCTCCACAGACAGATCAAGATGAGGAAGATTA----- 284

QY 176 lavalGlnProtheryrserAsnAsnGlnLutyrPheLeuIleValTyraSpValPheLysA 196

DB 285 -----AACGATTACAAACTCCGGAAGAAAGAACTTTTGAGG 321

QY 196 spValArgMetValPheAlaProPserSerValGlyLysPheGlyGlyAspThrAspA 216

DB 322 AT-----AACATAAGAAAATAAGAGACTGTGATTAGCA 354

QY 216 snTTPmet-----TTProArgHisThelGlyAspPheSerValPheArgValT 232

DB 355 ACTGGTAAATAATACGCAATAG----- 377

QY 232 yValAGlylAlaspaSnArgProAlaGluTYrSerLysAspaSnLysProTYrLysProV 252

DB 378 -----GAGGAAAGTCTAAAGAAATTCAAGAGGCTCGATCCA 414

QY 252 aLYrPheAlaAlaValsermetGlnGlyTYrLysAlaAspPYrAlaMetThrIleG 272

DB 415 TATACAGCGCTCTTATGATGTC-----GACTAC----- 443

QY 272 lYpHeProGlySerThraspaArgTYrLeuThrSerTTPoLYalGluAspaArgIleGlu 292

DB 444 -----CGAAATATTATCCCTCGTGAATATGCAAGAAATGCA 483

QY 292 sngLusnAsnProArgIleGluValArgGlylYleYsglnGlylIetTPlYsGluAlaM 312

DB 484 TGAAGAACCGCCAGTCAACCATGCCGA-----AACATCTGGACCGCTGCCA 531

QY 312 eTserAlaAspGlnAlarhrArgIleLysTYrAlaSerLysTYrAlaGlnSerAlaAsnT 332

DB 532 TC-----ACAACCTCCACGCGTCAACCAAGTCTGTGTAAGTACACAT 576

QY 332 yTTPrLysAsnserIleGlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgL 352

DB 577 ACATGAGAGATGTGGGC-----AACGTTCGTGTGGCC 612

QY 352 yAsrGlaGluGluArgAlaPheAlaAspTYrIleArgLysAsnGlyLysSerAlaValT 372

DB 613 GCCAGCTGTTCAGCC-----TGATGATATGCGACGCTGAGACAGCCCT 660

QY 372 yrgLYAspValleuserSerLeuGluLysAlaTYrLysGluGlyAlaLysAlaAsn---- 390

DB 661 GGCATTCTTACATC---AACTTCGAGCTGAGATACAAAGAGTGGAGCGGCGCCAGCA 717

QY 391 ----ArgLumethrTYrYrLeuSerGluThrLeuPheGlyGlyThrgGluValArgP 409

DB 718 TCTATGACGCTTGTGCTGCTGCTGACCCCTGCTGTAATAAATGATCAAGTACGCAAGAT 777

QY 409 he-----AlaGlnPheAlaAsnAlaLeuAlarhr----- 418

DB 778 TTGAAGAGAGCATGCTTACTTGTCTCATGCGCGGAATCTACAGAGCGGTGAGT 837

QY 419 -----AsnProAspAlaHisAlaGlyLe---LeuLysSerLeuAspAspI 433

DB 838 TTTTGGGATGAGCATATGAGACGAACACCTCATGTGCGCTTGTCTAAATTTGAGAGAA 897

QY 433 ySTYrLysAspTYrLeuProserLeuAspArgLysValLeuProAlaMetLeuAspIleV 453

DB 898 ACCAGAAAGATTT-----GAAAGGTACGAGTTATCTPACAAATATGCCCTGGAT---- 947

QY 453 aAlarGArgIleProAlaAspLysLeuProAspIlePheLysAsn-----ValIleA 471

DB 948 -----AGANTTTCAAAACAAGAGCCCAAGAACTCTTTTAAACATATACATCTTTG 999

QY 471 sPLysLysPheLysGlyAspThrLysLysTYrAlaAspPheValPheAspLysSerValY 491

DB 1000 AGAAGAAATTT---GGTACCGGAGGGGTATAGAAACATCATCTCGACCAACGAGGT 1056

QY 491 aLProTYrSerAspLysPheHisAlaMetLeuLyssermetLysPlys---GluLysPhea 510

DB 1057 TCCAGTATAGAGAAAGTGAAGGCTTAATCCACACATATATATGATGATGATTAT 1116

QY 510 lAlYsAlaIleGluLysAspProAlaValAlGluLeuserLysSerValIleAlaAlaAla 530

DB 1117 TACGCTGTGTGAAAGTGTGCGGAAGCTGAC-----ACTGTGC 1155

QY 530 rGAlaIleGlnAlaAspAlaMetAlaAsnAlaTYrAlaIleGluLysGlyLysArgLeuP 550

DB 1156 GGAAGACTATATGAGAGCCATTTGGCAATGTGCGACCCATCCAGGAGAGAGACAC----- 1211

```

Oy 550 hepheAlaGlyLeuArgGluMetTyrProGlyArgAlaLeuProSerAspAlaAsnHet 570
Db 1212 -----TGAAGCGCTACATCTAC----- 1229
Oy 570 hMetArgMetSerTyrGlySerIleLeuGlyTyrGluProGlnAspGlyAlaTTPtyrA 590
Db 1230 -----CTTGGGTTA 1239
Oy 590 snTYrHisThrThrGlyLysGlyValLeuGluLysGlnAspProLysSerAspGluPhea 610
Db 1240 ACTATGCACTCTATGAAG-----CTGAAGCCAGAGATCCGAGACAGACAGACAG 1293
Oy 610 laValGlnGluAsnIleLeuAspLeuPheArgThrLysAsnTyr 624
Db 1294 TATATCAAGCGTCT--TTGAACTAATTCTCACAAAAGTTTC 1334

```

Search completed: October 18, 2002, 00:00:54
 Job time : 1734 secs